Collective cognition and decision-making in humans and fish

Dissertation

zur Erlangung des akademischen Grades
Doctor rerum agriculturarum (Dr. rer. agr.)

eingereicht an der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

von

Romain Jean Gilbert Clément, M. Sc.

Präsident der Humboldt-Universität zu Berlin: Prof. Dr. Jan-Hendrik Olbertz
Dekan der Lebenswissenschaftlichen Fakultät: Prof. Dr. Richard Lucius

Gutachter:
1. Prof. Dr. Jens Krause
2. Dr. Max Wolf
3. Dr. Richard James

Tag der mündlichen Prüfung: 12. April 2016

Contents

Summary

Group living is a widespread phenomenon. One of its assumed advantages is collective cognition, the ability of groups to solve cognitive problems that are beyond single individuals' abilities. In this thesis, I investigated whether decision-making improves with group size in both humans and fish, thus using the strengths of each system.

In humans, I tested individual performance in simple quantity estimation tasks and a more difficult sentence reconstruction task first alone and then as part of a group. My question was whether groups were able to improve not only on average individual decisions, but also to beat their best members. Indeed, when a given problem is recurrent or too complex for individuals, groups were able to outperform their best members in different contexts. Furthermore, I showed that in a simulated predation experiment, groups of humans decided to stay or to escape using quorum thresholds based on movement behaviour without verbal communication, as has been shown in other animals. This simple movement mechanism allowed individuals in groups to simultaneously increase true positives and decrease false positives.

In the guppy, a freshwater fish from Trinidad, I tested in their natural environment whether individuals' ability to distinguish between an edible and a non-edible food item increases with group size. My results indicate that guppies had better chances to identify the edible food item when part of bigger groups. By investigating several populations with different ecological backgrounds, in particular differing in predation levels, I found that, despite a lower sampling activity in high predation habitats, predation did not affect the improvement of decisions in groups.

Overall, this thesis contributes to an improved understanding of collective decision-making, showing that collective cognition can arise from various interaction rules, such as simple aggregation of individual estimates, visual observation of others' movements, and group discussion. Furthermore, having tested problems ranging from estimating a quantity, distinguishing between edible and non-edible food items, deciding to stay or escape, and reconstructing a distorted message, I showed that grouping with others is beneficial in many situations for both humans and fish. This may suggest that the underlying mechanisms of collective cognition are remarkably similar across species.

Zusammenfassung

Das Zusammenleben in Gruppen ist im Tierreich ein weit verbreitetes Phänomen. Einer der Vorteile des Gruppenlebens könnte die sogenannte „Schwarmintelligenz“ sein, das heißt die Fähigkeit von Gruppen kognitive Probleme zu lösen, die die Problemlösekompetenz einzelner Individuen übersteigt. In der vorliegenden Dissertation untersuchte ich, ob die Gruppengröße beim Menschen und bei Fischen mit einer verbesserten Entscheidungsfindung einhergeht.

Beim Menschen analysierte ich zunächst das Abschneiden von Einzelpersonen, die später als Teil einer Gruppe getestet wurden, in einfachen Einschätzungsaufgaben sowie komplizierteren Satz-Rekonstruktionstests. Meine Frage war, ob es Individuen in Gruppen gelingt bessere Entscheidungen zutreffen als das einem durchschnittlichen Individuum der Gruppe alleine möglich wäre und ob Gruppen sogar die Leistung ihres besten Mitglieds in den individuellen Tests überbieten könnten. Tatsächlich konnte ich zeigen, dass Gruppen die Leistung des besten Mitglieds übertreffen, wenn die Problemstellung für Einzelpersonen zu komplex ist oder sich häufig wiederholt. Weiterhin gelang mir zu zeigen, dass Gruppen von Menschen bei einer simulierten Prädationssituation, ähnlich wie es bereits für andere Tierarten beschrieben wurde, anhand von so genannten „Quorum“-Regeln durch non-verbale Kommunikation entscheiden, ob sie bleiben oder flüchten. Dabei dienen einfache Bewegungsmuster als Schlüsselreiz. Individuen einer Gruppe erhöhen durch diesen Mechanismus gleichzeitig ihre echt positiven und verringern ihre falsch positiven Entscheidungen.

Beim Guppy, einem Süßwasserfisch aus Trinidad, untersuchte ich in deren natürlichem Habitat, ob die Fähigkeit einzelner Individuen zwischen einer genießbaren und einer ungenießbaren Futterquelle zu unterscheiden, mit der Gruppengröße ansteigt. Meine Ergebnisse zeigen, dass Guppys mit größerer Wahrscheinlichkeit eine genießbare Futterquelle identifizierten, sobald sie Teil einer größeren Gruppe waren. Untersuchungen an verschiedenen Populationen, die sich vor allem bezüglich des jeweiligen Prädationsdrucks in ihrem Habitat unterschieden, ergaben weiterhin, dass sich, abgesehen von einer niedrigeren Sampling-Rate in Habitaten mit hohem Prädationsdruck, Prädation nicht auf die Qualität der Gruppenentscheidungen auswirkt.

Die vorliegende Arbeit trägt zu einem tieferen Verständnis von kollektiver Entscheidungsfindung bei. Kollektive Intelligenz entsteht aus verschiedensten

Interaktionsregeln, beispielsweise durch die simple Zusammenfassung individueller Vorhersagen, visuelle Observation der Bewegung anderer und Gruppendiskussionen. Meine Arbeit lässt den Schluss zu, dass bei Problemen wie der Schätzung einer Menge, der Unterscheidung zwischen einer genießbaren von einer ungenießbaren Futterquelle, der Entscheidung zum Bleiben oder Flüchten und der Rekonstruktion einer Nachricht, die Bildung einer Gruppe, sowohl bei Menschen als auch bei Fischen, von Vorteil ist. Diese auffallenden Ähnlichkeiten über Artgrenzen hinweg weisen möglicherweise auf sehr universelle Mechanismen kollektiver Intelligenz hin.

Acknowledgments

I am grateful to the Leibniz-Insitute for Freshwater Ecology and Inland Fisheries (IGB) that provided me with space and funding during my PhD. IGB was an excellent and stimulating place to do my PhD. I also thank the Association for the Study of Animal Behaviour for travel grants to conferences.

I thank my supervisor Jens Krause, who shared his enthusiasm for science and wildlife, and who taught me many lessons, in experimental design, scientific reasoning and writing, *und alles*. Additionally, he has been an excellent travel companion during field work under conditions that were not always easy.

I thank Ralf Kurvers for the many discussions during which I learnt a lot, in particular concerning advanced statistical analysis, and also for his always swift and thorough feedback on my writings. Ralf's contribution is indeed reflected in that he is a coauthor on most of the chapters in the thesis. I am also grateful to Max Wolf for his insightful contributions during conversations and on manuscripts.

I am indebted to Stefan Krause for his crucial role during the analysis of complicated datasets involving language processing and other advanced computational methods. Discussions with Dick James, Kate Laskowski, Richard Mann, Thomas Mehner, James Herbert-Read and Ashley Ward improved various parts of this thesis in many ways. I am indebted to David Bierbach for his helpful feedback on the summary and the German translation. I thank Alex Wilson for his sound advice during the early stages of my PhD, and Karoline Borner for regular exchanges during our PhD years.

I am grateful to Indar Ramnarine who greatly facilitated our work in Trinidad. I also want to thank Nikolaus von Engelhardt, Leif Engqvist and Fritz Trillmich for hosting the experiments in Bielefeld, and all the participants of the 2011 and 2012 student course "Basismodul Biologie" and its tutors. I thank Marcus Ebert, Philipp Beer, Knut Hinrichs, Dominik Jost, Ivan Rodriguez-Pinto, Sonja Smith, and Simon Stäblein for assistance with data collection.

Finally, I thank my family and friends at IGB for their support during the long process of writing a PhD thesis. I thank Carolina for her support during the last months of the PhD. Last, I thank Newton, quiet observer of my writing who helped me focussing during the last phase.

1. General introduction

1.1 Collective behaviour

Collective behaviour is a widespread phenomenon that has long puzzled observers. Despite costs such as competition and increased probability of disease transmission, the ubiquity of group living across many species reflects the many advantages that it provides. Such advantages include antipredator defence, finding a mate, conserving heat and water, and reducing the energetic costs of movement (Krause and Ruxton, 2002; Sumpter, 2010).

At the beginning of the twentieth century, the naturalist Edmund Selous invoked telepathic faculties to explain the ability of flocks to perform fast synchronous changes of directions (Couzin, 2009). Now, various models have been developed to help explain the underlying mechanisms that allow fish schools and bird flocks to execute the synchronized fast group movements and escape manoeuvres. Early models based on particle physics used simple rules such as attraction to other individuals, alignment with neighbours, and repulsion from individuals that are too close (Breder, 1954; Radakov, 1973). Some other models known as topological models are based on a fixed number of neighbours that interact with the focal individual, regardless of their distance, rather than the number of individuals present within a given distance of the focal individual (Ballerini et al., 2008). More recently appeared visual models (Strandburg-Peshkin et al., 2013), in which individuals interact with other individuals that appear in their visual field (and occupy an angular area on their retina that is bigger than a threshold value). These recent models account for a great flexibility allowing groups to be very compact or very loose while maintaining cohesion and effective information transfer between the members. Understanding how members interact in a group is important to appreciate how information flows between members and enables group decision-making.

Moving from the interaction mechanisms governing groups, another important body of literature deals with understanding the adaptive consequences of group living. The dilution effect and the confusion effect, which are respectively a decrease in the probability to get caught (proportional with the number of individuals in the group) and an increase in the difficulty for the predator to focus on a particular prey in the group

(Ioannou et al., 2012) contribute to a better protection against predators for individuals in groups.

Some advantages come from cooperation and division of labour, specialized individuals being more effective at their task as has been shown in group-hunting dolphins (Gazda et al., 2005), or more commonly known in social insects.

In addition to group advantages that are mostly due to an increase in physical power, being part of large groups can provide advantages at the cognitive level, enabling individuals in groups to make better decisions, such as choosing the best navigation route, or detecting a predator earlier and from a greater distance through collective vigilance, known as the many-eyes effect (Pulliam, 1973), as observed in taxa ranging from water skaters (Treherne and Foster, 1980) to birds (Kenward, 1978).

1.2 Collective decision-making

All organisms, throughout their lives, are faced with many decisions. Whether to forage alone or in a group? What to eat? When and where to go? Whom to mate with? Where to breed? Etc. Therefore a lot of animal behaviour research investigates questions related to choice and decision-making. Decision-making in humans regarding every aspect of their life has been contemplated by philosophers since the antiquity.

Collective decision-making in humans

Owing to the inherent sociality of humans and the importance of decision-making in domains such as politics, economics and law, collective decision-making has been widely studied in humans. Building on pioneering works by Condorcet (1785) and Galton (1907), group decision-making has been studied extensively in humans by psychologists and economists (Fernandez-Juricic et al., 2004; Kerr and Tindale, 2004; Laughlin et al., 2002, 2003; Sunstein, 2005, 2005; Surowiecki, 2004).

However, collective decision-making has been shown to have limits. Depending on the conditions, group decision can either increase or decrease the quality of the decision (Esser, 1998; Janis, 1971; Koriat, 2012; Lassila, 2008). Social influence for example, by reducing independence, can be detrimental as the bare knowledge of others' opinion may negatively affect the collective outcome, even without communication (Lorenz et al., 2011).

Recently, collective decision-making in humans has been studied a lot in behavioural research with an evolutionary or behavioural ecology focus (Dyer et al., 2008; King et al., 2011a, 2011b; Krause et al., 2010, 2011b; Kurvers et al., 2014a; Wolf et al., 2013).

Collective decision-making in (non-human) animals

Despite the numerous studies on group decision-making in humans and individual decision-making in animals, collective decision-making has long been neglected in animals and the number of studies investigating collective decision-making in non-human animals increased sharply only recently.

In many situations, conflicts of interest can arise (for instance, the optimal choice for an individual can vary within a group depending on its age, sex, nutritional state, etc). In order to maintain cohesion, animals that live in groups need to make consensual decisions. But, except from social insects, most animal groups are heterogeneous and it is unlikely that all individuals have the same needs at the same time (Conradt, 2011, 2012; Conradt and List, 2009, 2009; Conradt and Roper, 2003, 2005, 2007, 2009, 2010; Conradt et al., 2009).

Collective decision-making can result in benefits (increased vigilance through many eyes, faster decision, increased navigational accuracy, self-organized orientation to local neighbourhood) but can also present several costs (increased decision time, competition, disease transmission, cascade of information). This thesis deals with situations where it is in the interest of all members to take the best decision, such as escaping from a predator, or deciphering a message. Only chapter 3 may involve some level of competition between the members of a group, but all have the same interest in making the correct choice.

Collective cognition

Groups are able to achieve things that individuals cannot and improved decision-making is at the root of several of the advantages offered by group-living. For example, water skaters, through the many eyes effect, are able to detect predators from a greater distance. The many eyes effect is an example of collective cognition and leads to an increase of overall vigilance for the group, while allowing a decrease of vigilance at the individual level (Treherne and Foster, 1980).

Collective cognition is a particular case of collective behaviour that has its origins in the fields of complexity and self-organization. It is defined as a process by which "two or more individuals independently collect information that is processed through social interaction and provides a solution to a cognitive problem that is not available to single individuals" (Krause et al., 2010). In short, collective cognition is the ability that groups have to solve cognitive problems that are beyond individual capabilities. Depending on the field in which it is used, or the organisms studied, collective cognition is also known as collective intelligence when used generally on animals (Couzin, 2009), swarm intelligence when applied more specifically to social insects or algorithms (Bonabeau et al., 1999), the Wisdom of Crowds (Surowiecki, 2004), group decision-making or team decision-making in psychology (Kerr and Tindale, 2004; Sunstein, 2005). The expression Swarm Intelligence was first coined by Beni & Wang (1989), to describe "systems of non-intelligent robots exhibiting collectively intelligent behaviour evident in the ability to unpredictably produce 'specific' (i.e. not in a statistical sense) ordered patterns of matter in the external environment". The term "collective intelligence" was first applied to biological systems by Franks (1989) to describe the ability of ant colonies to solve specific problems that appear out of reach for individual ants. It has since then been applied not only to other social animals, ranging from bees (Garnier et al., 2007) to humans (Krause et al., 2011b), and has even been proposed as a coordination mechanism in complex plants roots systems (Baluska et al., 2010).

Indeed, it has been shown that ants are able to choose the shortest route between a food source and their nest using trail pheromone (Goss et al., 1989). Using simple quorum rules, accurate decisions are achieved by bees (Seeley et al., 2006; Visscher and Camazine, 1999) and ant colonies (Pratt et al., 2002) that are forced to find a new nest after being evicted. It has also been shown that speed and accuracy augment with group size in fish shoals avoiding a predator model (Sumpter et al., 2008a; Ward et al., 2008, 2011)

Collective cognition can be achieved through different mechanisms. In addition to benefiting from the possibility to follow a clear leader (Couzin et al., 2005; Danchin et al., 2004; Reebs, 2000), being part of the group can also improve decision-making even when nobody has clear information (Couzin et al., 2011). For example it has been demonstrated that migrating groups use averaging in order to navigate more accurately and reach their destination (Codling et al., 2007; Faria et al., 2009; Hancock et al., 2006; Wallraff, 1978, 2001). Additionally, group members are still able to distrust a single bad

leader leading them close to a dangerous area and need a critical number (Ward et al., 2011).

In a group, information is rarely equally distributed among all individuals. Individuals may, for example, differ in their proximity to a predator, or in previous experience about a food location or a past migration route. Therefore leadership can arise in non-hierarchical groups from differences in information. Couzin et al. (2005) predicted that a minority of informed members as small as 5 to 10 % is necessary to lead a whole group towards the right location. This has been empirically verified by Dyer et al. (2008, 2009) using human volunteers. Moreover, recent studies have shown that the presence of uninformed individuals in a group not only favours majority decisions by counterbalancing despotic tendencies of minorities (Couzin et al., 2011), but also improve the stability of the decision-making (Leonard et al., 2011).

1.3 Outline of the thesis

The main aim of my thesis was to investigate whether decision-making improves with group size. For this, I compared individual performance with group performance using an array of simulation experiments on humans (**chapters 2, 4, 5**) and a field study on Trinidadian guppies, *Poecilia reticulata* (**chapter 3**). Using two different systems allowed me to investigate collective cognition from different angles, each system presenting its own strengths.

While investigating whether group decisions improve compared to decisions of individuals, I was also interested in finding out whether collective decisions can also be better than decisions of the best individuals. This is important in order to understand whether top performers join groups because of cognitive benefits or because of benefits related to other group advantages such as, for example, dilution effect. I tackled this question in **chapter 2** where I studied groups' potential performance at solving a recurring task, and in **chapter 5**, where I studied groups' performance at solving a more complex task. Studying humans allowed me to know someone's private information before they acted on it, and before they shared it with other members of the group. This allowed delineating private information from social information, which can be difficult and time-consuming in animals (because this requires individual testing; therefore, most animal studies on collective cognition compared individuals with groups made of different individuals that have not been tested as singletons, making it difficult to assess

to which extent particular individuals improve when part of a group). Still taking advantage of the possibility to know individuals' decision before and after interaction with others, in **chapter 4** I studied the ability of groups to improve decisions using a design that simulated natural situations and that allowed individuals to interact via movement, mimicking the decision process observed in wild animal groups.

Finally, I wanted to see whether the cognitive benefits observed in human groups apply to animals living in natural conditions. Most fundamentally, I was interested in exploring ecological selective pressures that may have influenced the evolution of collective cognition in wild animals. This is possible by comparing different species or different populations and trying to understand how differences in behaviour reflect differences in ecology (Davies et al., 2012). In **chapter 3**, I used the comparative method to investigate the effect of predation on collective cognition in Trinidadian guppies. I compared the decision improvements in groups across four populations that differed in predation pressure, i.e. two with a high predation level, and two with a low predation level. Furthermore, in this experiment, the group sizes used occurred naturally and were not artificially made by separating already existing groups or putting together unfamiliar individuals.

I chose to present the chapters of my thesis following the increasing complexity of the interactions between group members, which also reflects the increasing level of complexity of the tasks that they had to solve.

Chapters 2 and 3 used a simple assessment of collective cognition based on decision accuracy.

It is known that groups generally outperform average individuals at estimation tasks. But they are often beaten by one or a few individuals whose estimates are even closer to the real value. **Chapter 2** explored the advantage of collective cognition compared to top performing individuals in solving a task that is cognitively simple, but repeatedly encountered over time. In this study, individuals were asked to estimate the number of dots that appear on a screen. This estimation task was carried out repeatedly to find out how often individuals beat the group. I compared performance of the best individuals to group performance obtained from a combination of the estimates given by individual members. Depending on the problem that is encountered, the incentive to join the group may be strong enough even for the best individual. For example, in the case of

predator detection, missing to spot a predator only once can be fatal and joining the group is beneficial if such a problem is encountered frequently.

Chapter 3 extended collective cognition to animal groups in the wild. Testing the performance of individuals in various group sizes, I explored the ability of guppies to distinguish two stimuli (one edible and the other not) that are visually very similar. This experiment was run across several populations to investigate the effect of predation pressure on decision-making abilities, and in particular differences in the use of private and social information.

Chapters 4 and 5 took advantage of humans as a study system to investigate collective cognition to solve more complex tasks and a possibility to look into the collective decision mechanism at a finer level.

In **chapter 4**, groups of humans had to repeatedly distinguish between two cryptic images in a simulated predation detection experiment in which they were only allowed to indicate their preference via movement, in a way that is similar to group decision-making observed in some wild animal groups (such as in mammal herds, bird flocks, fish schools).

It is generally difficult to investigate complex problems as the correct solutions are not always obvious and most of the previous work on swarm intelligence has been carried out on simple estimation tasks. In **chapter 5**, I explored collective cognition as a tool for solving more complicated problems. After listening to a distorted and hardly understandable announcement, participants were asked to reconstruct the original message individually or as a group, by discussing.

2. Collective cognition potential in humans: Groups can outperform high-performing individuals

Clément RJG, Krause S, Faria JJ & Krause J

The possibility that individuals in animal groups, including humans, can make collective decisions that solve cognitive problems that single individuals cannot solve, or not in the same way, has attracted much attention in recent years in the context of collective cognition (CC). A common problem when comparing the problem-solving ability of groups and single individuals is that many studies only looked at one-off performances which makes it difficult to distinguish between individuals of high ability and those that made a lucky guess. Here we examined performance profiles of individuals regarding a repeated quantity estimation task which demonstrated that there was significant variation in cognitive ability within a human population. For a single estimate, 13.4% of individuals could beat the group performance. However, if repeated estimates were taken into account, then group performance was superior to even the best individual performers in the group after a sequence of at least 8 estimates. This result suggests that for certain cognitive problems that are encountered repeatedly, joining groups may always be advantageous, even for individuals of high cognitive ability. We discuss our results in the context of the evolution of collective cognition.

Unpublished manuscript

Introduction

One of the benefits of group living in animals comes from improved decision-making resulting from independent information collection and processing among group members (Krause et al., 2010). The fact that individuals in groups can solve cognitive problems in a way that cannot be implemented by single individuals – a process known as collective cognition – has attracted a large amount of interest in several areas (biology: Couzin, 2009; Krause et al., 2010, psychology: Kerr and Tindale, 2004; Sunstein, 2005; sociology: Mathieu et al., 2008; economics: Armstrong, 2001; Wolfers and Zitzewitz, 2004).

In humans, it has been shown that, for many kinds of problems, the group's collective performance is better than that of an average individual's (Hong and Page, 2004; Sunstein, 2005). This phenomenon, known as the "wisdom of crowds" or "many-wrongs principle" was pointed out as early as 1907 by Galton and seems to play an important role in some group formations such as migrating animals (Bergman and Donner, 1964; Hamilton, 1967; Simons, 2004; Wallraff, 1978) and several models support this hypothesis (Codling et al., 2007; Grünbaum, 1998; Hancock et al., 2006). These models are confirmed by Dell'Ariccia et al. (2008), who showed that homing pigeons travel more efficiently in groups than alone and reach their destination faster. However, Biro et al. (2006) showed that if the disagreement regarding the route is too high, either the group would split or one pigeon would become the leader. In humans, when the information is unevenly distributed among the members of a group, knowledgeable individuals can still lead the group to the right destination, even without communication between members (Dyer et al., 2008, 2009). But when nobody owns enough information, the wisdom of crowds plays an important role in reaching the correct destination when the group size is large enough and when uncertainty is high (Faria et al., 2009). Dyer et al. (2008, 2009) showed that even when the information is not evenly distributed and only a minority is able to lead, groups of humans still reach their destination, even without communication. Some individuals tend to overestimate while others underestimate and groups usually outperform individuals because averaging these estimations leads to a reduction of the error component (Fischer and Harvey, 1999; Simons, 2004). When an objective and demonstrable answer does exist such as estimating the temperature of a room, the number of beans in a jar or the ranking of different weights (Sunstein, 2005 and references therein), the group's mean or median answer often comes

very close to the real value and is better than the vast majority of individual estimates. However, it has been pointed out that the group performance usually did not surpass the best individual in a given group and that group processing can actually lead to gains as well as losses (Kerr and Tindale, 2004). Very few studies actually reported groups' outcomes that were better than the performance of every member (Kerr and Tindale, 2004). It is generally considered a truism that groups do better than single individuals, since most statistics is based on the assumption that a larger sample size is more likely to return an average value closer to the true mean. But this is not always the case. For example, Krause et al. (2011) showed that in some cases, as sample size increases, the mean becomes worse, and when it comes to decision-making, benefits from grouping behaviour are not always high enough to outcompete the best individuals.

Furthermore, interactions between members of the group do not always improve the group's outcome because of inherent obstacles such as peer pressure or informational influence (Lorenz et al., 2011; Sunstein, 2005). Also, Krause et al. (2010) showed that, in some cases, groups can even do worse than some particular individuals, for example while asked to solve problems in which the answer is highly counterintuitive. Therefore, rather than testing swarm intelligence per se, in this experiment we show its potential for solving this kind of problem, by removing any direct interaction and considering a repeatable situation where individual differences in the ability to solving the problem are measurable.

Group performance has been studied a lot, particularly in humans (Kerr and Tindale, 2004; Laughlin et al., 2002, 2006), but a problem with these previous studies is that only a single event was considered. It was therefore not possible to tell whether the individual who got the best result and beat the group performance was an "expert" at doing the task or was just being "lucky". Experts are characterised by a consistently high performance (Shanteau et al., 2002). Our study differs from the previous ones by establishing a performance profile over several trials for each individual, enabling us to compare the performance of each individual against each other and against groups of different sizes.

A recent model predicted that in single-shot decisions, experts are almost always more accurate than the collective across a range of conditions, but that for repeated decisions – where individuals are able to consider the success of previous decision outcomes – the collective's aggregated information is almost always superior (Katsikopoulos and King, 2010). Our study tested these predictions. We presented

students with a set of 10 assessment tasks to quantify differences in personal ability. To simulate a kind of problem that could be encountered by an individual in reality, we used a simple quantity estimation; i.e. estimating the number of dots on a screen. We predict that when a task is repeated several times, the group eventually outperforms even its best members.

Methods

Data collection

The experiment was carried out at the University of Applied Sciences Lübeck, Germany, in November 2008 and October 2009. All data were collected anonymously and with the permission of the participants. Students from different technical disciplines were asked to estimate the numbers of dots on 10 pictures each of which was shown on a projection screen for 10 seconds. The screen size was 160 × 120 cm and the diameter of each dot was 1.2 cm. After a picture with dots was shown the participants had 15 seconds to write down their estimate before the next picture was shown. During the time intervals a uniformly grey picture was shown.

All participants saw the same pictures in the same order. Each picture was constructed by first picking a random number n in the range 50 … 500 and then randomly placing n dots in the visible area of the screen without overlapping. To make the estimations more difficult, the dots were not uniformly placed but the probability of a position decreased with increasing distance from the centre of the screen (inset of Fig. 2.1). This prevented participants from simply counting the dots in a small subset of the area and then scaling up to obtain the total number. From the 98 answer sheets one was excluded from the evaluation because of extremely high and deliberate-looking errors.

Analysis

Each picture contained a different number of dots. In order to make errors comparable between pictures we used relative errors to measure performance. More precisely, for each individual i and picture k we defined the individual relative error $d(i,k) = | g(i,k) - c_k | / c_k$, where $g(i,k)$ was the guess of individual i for picture k, and c_k was the correct value for picture k. In the same way we defined the collective relative error for picture k as $D(k) = | \mathrm{mean}_{i \in I}\, g(i,k) - c_k | / c_k$, where I was the set of all individuals ($N = 97$) and $\mathrm{mean}_{i \in I}\, g(i,k)$ was the mean of the guesses of all individuals in the set I for picture k.

Relative error as a function of the correct value

Do the $d(i,k)$ and $D(k)$ depend on the magnitude of the correct values c_k? It might be the case that large numbers of objects are more difficult to estimate than small ones even in terms of relative errors. This phenomenon is known as the Weber-Fechner law (Dehaene, 2003; Nieder and Dehaene, 2009). If so, the question arises whether this issue only affects the individual relative errors, or both the individual and collective relative errors. To investigate these questions we computed the mean individual relative error $mean_{i \in I}\, d(i,k)$ for each picture k. Then we conducted regressions with the correct value c_k as independent variable and $mean_{i \in I}\, d(i,k)$ and $D(k)$ as dependent variables.

Differences in individual ability

In order to find out whether some individuals consistently performed better than others we split the sequences of guesses in two halves, the guesses for pictures 1-5, and the guesses for pictures 6-10. For each individual i we computed the sums of relative errors for both sub-sequences, $\Sigma_{k=1...5}\, d(i,k)$ and $\Sigma_{k=6...10}\, d(i,k)$, and determined their correlation. A significant correlation would be indicative of individual consistency. The $d(i,k)$ tended to increase with increasing correct value c_k (see results section). Therefore, it seemed problematic to regard these error sums as measurements at interval scale or ratio scale level. However, the error sums certainly provide a measurement at an ordinal scale level and it made sense to use a rank correlation coefficient. Here we used Kendall's τ, mainly because it has a simple and intuitive interpretation.

Individual performance vs. collective performance

We compared individual and collective performance across (subsets of) the 10 pictures by comparing their cumulative errors. We computed the cumulative error for a subset S of the 10 pictures by summing up the errors for all pictures in S. More precisely, the cumulative error of individual i on a subset S of pictures was defined by $\Sigma_{k \in S}\, d(i,k)$, and the cumulative collective error on S was $\Sigma_{k \in S}\, D(k)$. For example, if S contained the pictures 1, 4, and 5, then the cumulative error of individual i on S was $d(i,1) + d(i,4) + d(i,5)$ and the cumulative collective error on S was $D(1) + D(4) + D(5)$.

For most single pictures some individuals outperformed the group (i.e. $d(i,k) < D(k)$ for some individuals i and pictures k). On average, 13 individuals had a smaller error than the group for a given picture. However, regarding the cumulative error on the set of all 10 pictures no individual was better than the group. The question arose about what happens in between these extreme scenarios for intermediate numbers of estimates. To answer this question we compared the individual and the collective cumulative errors on subsets of pictures. For each subset we counted the number of individuals with smaller cumulative error than the group. In order to determine if and how this number decreases with increasing size of the subset we did this for all 2^{10} - 1 non-empty subsets of pictures.

Additionally, to assess the influence of differences in individual abilities on the number of individuals that perform better than the group on subsets of pictures, we randomised the guesses across the individuals. More precisely, for each picture we permuted all guesses and performed the above described computations for all non-empty subsets of pictures. We repeated these steps 100,000 times to approximate the probability distribution of the number of individuals that outperform the group for each size of subsets of pictures in a null model where all individuals have the same abilities.

Statistical analyses were performed using R version 2.10.1.

Results

Relative error as a function of the correct value

The mean individual relative error increased with the magnitude of the correct value c_k (Linear regression: $F_{1,8} = 21.15$, $P = 0.002$, $R^2 = 0.726$; Fig. 2.1), while the collective relative error did not increase ($F_{1,8} = 0.41$, $P = 0.54$, $R^2 = 0.048$; Fig. 2.1). The individual relative errors increased linearly (rather than exponentially as would be expected if following a Weber-Fechner law; Fig. 2.1). The collective relative error did not increase probably because the ratio of guesses that underestimated to guesses that overestimated the correct values was roughly the same regardless of the correct value.

Differences in individual ability

Individual performance on the two halves (two sets of 5) of pictures was significantly correlated (Kendall's tau: $\tau = 0.43$, $z_\tau = 6.30$, $P < 0.001$) indicating that there were strong inter-individual differences in estimation performance. From the τ value we could

conclude that an individual i_1 that performed better than an individual i_2 on the first half of 5 pictures would also perform better than i_2 on the second half of 5 pictures with a probability of approximately 72%.

Individual performance vs. collective performance

The cumulative collective error on the set of all pictures was 1.08 whereas the cumulative errors of individuals were much larger and ranged from 1.73 (best performer among the 97 individuals) to 22.36. The evaluation of observed cumulative errors on subsets of pictures showed that the number of individuals that performed better than the group decreased exponentially with increasing size of the subset showing that even the strongest individual performer could not beat the group after 8 attempts (Fig. 2.2). The mean numbers extracted from the null model also decreased exponentially but were significantly smaller than the observed ones as can be seen from the 95% confidence interval (Fig. 2.2).

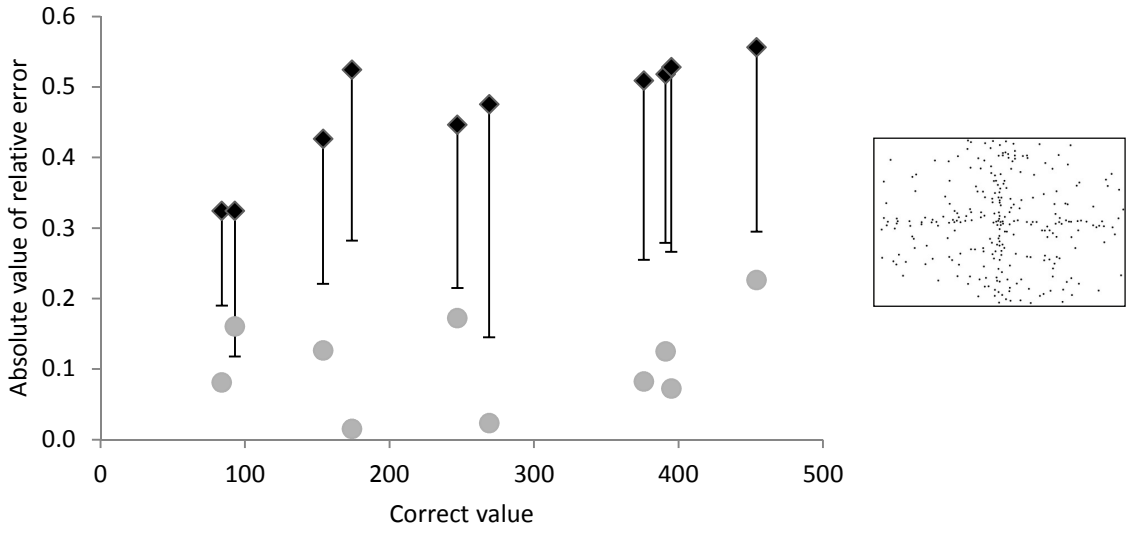

Figure 2.1: Mean individual error (black squares) and collective error (grey circles) on single pictures as a function of the correct value. The black bars indicate the interval that covers approximately 66% of individual errors that are smaller than the mean. We used this range rather than the standard error because of the highly asymmetric nature of the distributions. Inset: Example of the pictures shown to the participants.

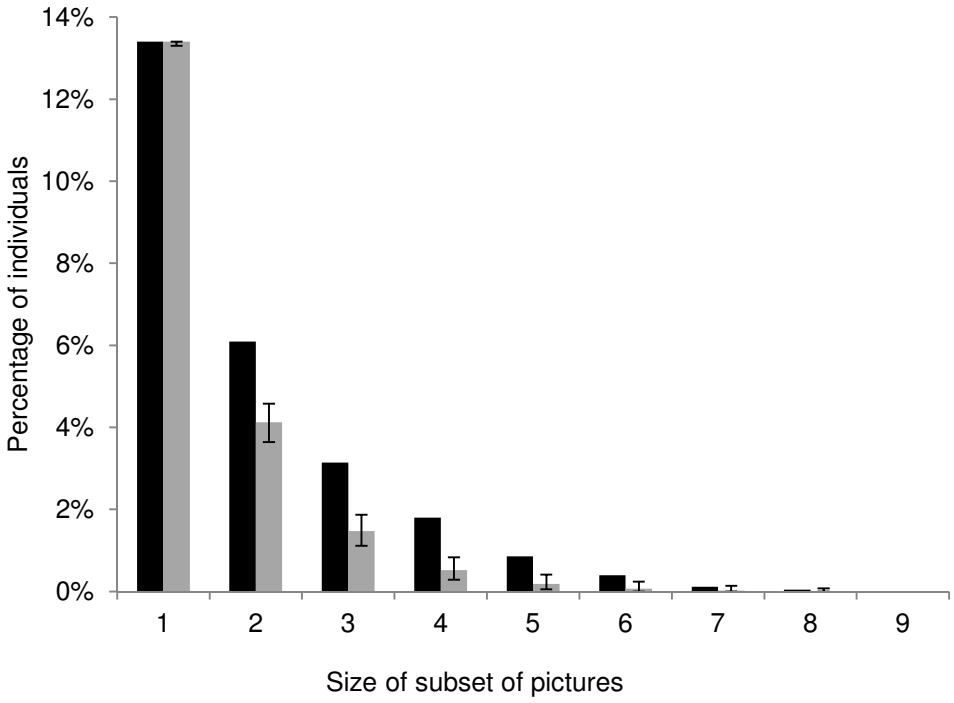

Figure 2.2: Percentage of individuals that outperform the group as a function of the size of the subset of pictures (black bars). After 8 guesses no individual in the data set was capable of beating the collective guess. For comparison the performance of the group is shown after removing individual performance differences from the data set (grey bars with 95% confidence intervals).

Discussion

Our results show that, although individual relative error increases as the correct value increases, the collective relative error remains constant. One novel aspect of this study lies in the use of repeated trials by the individuals, allowing us to establish individual profiles, which showed a significant inter-individual variation. But despite this variation, the group performance beats even the best individual after 8 guesses. These results are a strong indication that the swarm intelligence potential of the group becomes particularly important when the same, or at least similar, cognitive problems are encountered more frequently. In contrast 13.4 % of individuals could beat the group for a single guess.

This result suggests that experiments in which groups were not able to beat the best individual members (Kerr and Tindale, 2004; Mathieu et al., 2008; Sunstein, 2005) used sample sizes that were either too small to provide a good collective guess or the lack of repeated trials may have given a false impression of individual quality. There are, however studies which found that even relatively small groups (of 3-5 members) can outperform their best members for some kinds of problems (Clément et al., 2013; Laughlin et al., 2002, 2006). These are tasks in which groups achieve qualitatively better solutions to complex problems where a connection between letters and numbers need to be made and which required the development of equations. The mechanism, however, by which groups achieved this superior performance remains largely unclear. We conclude that simple estimation tasks (like the one investigated in this paper) require repeat performances to reliably assess the performance level of groups and single individuals whereas highly complex cognitive tasks can already show up a performance difference between groups and the best individuals in single trials.

What do these results tell us about the evolution of grouping via collective cognition? The results of our study indicate that the CC-benefit of group membership accumulates over time with repeat performances. Laughlin et al. (2002) suggests that an instant benefit of making even just a single decision in a relatively small group can be obtained in the case of complex problems. It would be interesting to see whether Laughlin's approach can be modified for use with species that are known for their cognitive abilities such as non-human primates or cetaceans.

Our results indicated in our scenario how often a cognitive problem needed to occur for the group to achieve higher performance levels than the best individual. But another important consideration is of course the fitness benefit that would be associated

with this higher performance level. This benefit would depend on how important the cognitive problem in question is and how frequently the individuals encounter this type of problem. If the problem occurs very frequently, then even a small fitness benefit may add up over the life-time of the animals (Davies et al., 2012). However, if the problem occurs only rarely, then we should expect it to be an important one if it is to lead to the evolution of grouping via CC. Finally we also need to consider the potential costs of grouping, such as competition (Krause and Ruxton, 2002), relative to any benefits before we can predict whether factors such as CC could lead to the evolution of grouping.

Research studies on cognitive abilities in animals and humans show strong evidence for inter-individual variation (Bell et al., 2009; Dall et al., 2004; Deary et al., 2010; Healy et al., 2009; Sih et al., 2004a, 2004b). However, it remains a considerable challenge to predict the collective performance of groups that are composed of individuals of different cognitive abilities. Different outcomes are possible depending on the group composition and the information processing rules (Krause et al., 2010). This is a promising avenue for future research that takes personality differences in animals into account when looking at collective decision-making and swarm intelligence in particular.

If all individuals greatly benefit from overcoming their cognitive limitations through CC in groups, mechanisms that favour group formation and maintenance are likely to have evolved. However, what about a population in which a few (very good) individuals might not benefit greatly from being in a group (regarding decision-making) whereas the majority would? If some individuals are superior decision-makers, they might do better staying mostly alone and thereby reduce competition costs. For example, regarding more traditional grouping benefits, if living in groups has evolved because of the many-eyes effect (Krause and Ruxton, 2002), and one individual is “super aware”, then it might be expected to spend less time in groups because the costs (competition for example) are likely to outweigh the benefits (in terms of spotting predators). In the context of cognitive problem solving, we could potentially have the emergence of two strategies such as low cognitive performers that are highly gregarious and high cognitive performers that are largely solitary, provided that they have equal fitness. Potentially there could even be a strategy continuum whereby individuals of different cognitive abilities use CC to different degrees. It would then be interesting to test whether individuals with lower cognitive abilities demonstrate higher grouping tendency than individuals with higher abilities and if this relationship is affected by the type of problem encountered. Another important consideration in this context is that the development and

maintenance of neural tissues is very expensive (Laughlin et al., 1998). Therefore swarm intelligence might often be the energetically cheapest option to solve cognitive problems.

In this experiment, the individuals did not exchange any information while making the decision. In humans, direct interactions are not always required for swarm intelligence because, once we know how the information should be processed, we can use a computer algorithm to do the job (Krause et al., 2010; Wolfers and Zitzewitz, 2004). In fact there is some indication in the psychological literature that real interactions between individuals can be detrimental to the swarm intelligence potential of groups because of communication barriers between them (Mathieu et al., 2008; Sunstein, 2005). Therefore, at least for this range of problems, aggregating the estimates from each individual instead of allowing the members to communicate with each other is the safest solution to guarantee the quality of the decision.

One aspect we have neglected so far is the question of how the performance of the group depends on group size. Unfortunately this was beyond the scope of our study because we only had 97 individuals. To assess the performance potential of different group sizes one would need a dataset of many more individuals and their responses, to randomly draw groups of different sizes and investigate their performance relative to that of the best group member. Statistical textbooks (for example, Dorofeev and Grant, 2006) generally give the advice that a correction factor has to be taken into account when sample sizes are greater than 5% of the population. This suggests that the pool from which samples can safely be drawn for the above purpose needs to be at least 20 times larger than the largest sample size. Therefore, if we want to examine the performance of groups up to size 50, we would need a minimum of 1000 individuals in our study. Experimental work on sparrows, *Passer domesticus*, showed that larger groups were more successful in solving cognitive problems (Liker and Bókony, 2009). However, they compared groups of 6 birds with those containing 2 birds and did not include single animals. The latter would be valuable in terms of understanding the evolution of grouping via cognitive benefits in groups which remains an interesting field for future research.

3. Collective cognition in guppies: a cross-population comparison study in the wild

Clément RJG, Mann RP, Ward AJW, Kurvers RHJM, Ramnarine IW & Krause J

Collective cognition has received much attention in recent years but most of the empirical work has focused on the increase of decision-performance with group size in single populations. Here we investigated collective cognition in multiple populations that are subject to different ecological conditions. Guppies (*Poecilia reticulata*) were given a simultaneous choice between an edible and a non-edible stimulus. We quantified the response of fish to the test stimuli in various group sizes across four populations that differed in predation risk. Our results show that sampling activity was higher in low predation populations compared to high ones but not affected by group size or sex. Decision accuracy increased with group size and with sampling activity of the focal individual, i.e. individuals that sampled more, and individuals in larger groups had a higher approach and peck accuracy. Group size had a significant positive effect on the probability that the first approach was made towards the edible stimulus but not on the accuracy of the first peck. Our results suggest that the use of private information (direct personal sampling) and social information (observing other group members) is context-dependent in guppies and differences in predation regimes only influenced sampling activity but not decision accuracy.

Unpublished manuscript

Introduction

Sociality confers many advantages to animals, such as better anti-predator defence or spotting food (Krause and Ruxton, 2002). Many of these advantages result from collective cognition, which is the ability of members of groups to make decisions that are better than those made by single individuals (Krause et al., 2010).

Animals often gain advantage from paying attention to cues that are inadvertently displayed by their conspecifics, using this social information to complement their own information in order to make better decisions (Danchin et al., 2004). For example, rats infer from their congeners' breath whether some food is safe or not to consume (Galef, 1991), whereas starlings observe foraging success of their flockmates to assess the quality of a food patch (Templeton and Giraldeau, 1995). Social information is also used by social animals to make collective decisions based on quorum thresholds. For instance shoals of stickleback (*Gasterosteus aculeatus*) use quorum thresholds to evaluate predation risk (Ward et al., 2008) or to locate foraging patches (Ward et al., 2012). Similarly, ant colonies that are forced to emigrate from their nest are able to choose the best available new nest by accelerating the recruitment process once a threshold number of ants have made a decision in favour of a particular location (Pratt et al., 2002; Sumpter and Pratt, 2009). It was also shown in fish shoals that the speed and the accuracy of decisions increased with group size (Ward et al., 2011).

One important selection pressure affecting the use of social information is thought to be predation risk (Devereux et al., 2006; Elgar, 1989). For example, experimentally increasing the perceived risk of predation in the lab resulted in an increased reliance on social information in Minnows (*Phoxinus phoxinus*) due to increased cost of gathering private information (Webster and Laland, 2008). Three-spined sticklebacks (*Gasterosteus aculeatus*) have a thick body armour making it less risky to collect private information, as compared to the sympatrically occurring nine-spined sticklebacks (*Pungitius pungitius*), which lack this armour and are more vulnerable to predation and more prone to using social information (van Bergen et al., 2004; Coolen et al., 2003, 2005). Furthermore, it has been shown that predation risk can influence social structure of animal groups (Kelley et al., 2011), possibly affecting information transfer, and therefore social learning. However, a largely unresolved question is how individuals from populations under different predation regimes in the wild differ in their use of private versus social information.

Many fish species feed on objects falling into the water, and individuals have to respond quickly to a stimulus and consume it before others do (Krause, 1993). However, not every fallen object is edible and the fish will often approach objects that they cannot consume at the cost of wasting energy, missing simultaneously occurring and genuine feeding opportunities, and exposing themselves to predation (Hall et al., 2013; Ioannou and Krause, 2009).

In many natural situations, timing matters, and little differences can lead to different outcomes among conspecifics. For example, detecting and responding to a predator earlier increases the chance of survival, while predators focus on individuals that react more slowly (Kenward, 1978; Post et al., 2013). Similarly, in a foraging context, being part of a group implies competing with other members when the food is limited and a rapid response to opportunities can increase food intake. For example one might benefit from social information about a fallen object by watching how the others react to it, but a delayed response may mean a missed opportunity. Authors previously looked at single populations in the context of decision accuracy (Sumpter et al., 2008a; Ward et al., 2008, 2012). Here we used the comparative approach to investigate groups of free-ranging guppies from four different populations facing different predation regimes to study the effects of predation risk on the evolution of collective cognition.

We studied the effects of group size and predation level on decision accuracy by presenting fish simultaneously with an edible and a (similar looking) non-edible item. We quantified how frequently the edible and the non-edible stimuli were targeted as a measure of decision accuracy. Furthermore, we quantified general sampling activity of individuals as a proxy of their level of private information. We studied how group size and predation level affected this sampling activity, and how sampling activity, in turn, affected decision accuracy.

Carotenoid pigments have been shown to confer health benefits (Kolluru et al., 2006). They are sequestered for use in courtship displays by male guppies (Kodric-Brown, 1989), thus also conferring fitness benefits. They cannot be synthesised by guppies and must be obtained through their diet (Fox, 1976). They are present in orange and red fruits that are abundant in the rainforest and often end up in the streambed, where they are very attractive to guppies, both males and females (Rodd et al., 2002). However, at the same time the streambed usually contains many other reddish objects including pebbles that are non-edible. We therefore presented edible and non-edible stimuli made

out of red materials that resemble each other to single fish and shoals of various sizes and evaluated the decision accuracy as a function of group size.

Due to the anti-predator effects of group-living, sampling is usually safer for individuals in larger groups (Krause and Ruxton, 2002). We therefore predicted that i) sampling activity would be positively correlated with group size. Given that sampling should be safer in lower predation sites we also predicted that ii) sampling activity is greater in low predation sites than in high predation sites. Finally, we expected that iii) decision accuracy would be positively correlated with group size, as predicted by previous work (Ward et al., 2008, 2011).

Methods

Experimental setup

The study took place in the Turure River (lower: 10°39'27"N, 61°9'48"W; upper: 10°41'7"N, 61°10'23"W) and the Aripo River (lower: 10°39'1"N, 61°13'26"W; upper: 10°40'55"N, 61°13'51"W) in Trinidad (March 2013, 2014 & 2015). These rivers consist of interconnected pools inhabited by different-sized guppy populations (*Poecilia reticulata*). Both streams are known to have a sharp gradient in predation pressure: in the lower sections (below the main falls) characids and cichlids are present, which heavily predate on guppies. These predators are absent in the upper sections (above the main falls) (Magurran, 2005). This provides a unique opportunity to use a natural gradient in predation pressure to investigate its effect on collective cognition. We sampled populations from below and above the main falls in both rivers.

Individuals from our target populations are known to respond strongly to orange and red items falling on the water surface (usually these are edible fruits dropping from trees into the stream). Therefore, an edible stimulus and a non-edible stimulus were made out of red material. The edible stimulus was a piece of red bell pepper and the non-edible stimulus was cut out of red plastic. Both stimuli had the same shape (9*5*2 mm) and each was fixed to a weighted monofilament line (40 cm long, ø 0.2 mm) attached to the end of a wooden rod, 20 cm apart from each other. We first conducted pilot trials to verify that both stimuli were effective with fish in the wild when presented in isolation (mean ± SE number of approaches towards the edible stimulus: 4.1 ± 0.5 (n = 30), towards the non-edible stimulus: 2.6 ± 0.4 (n = 37); mean ± SE number of pecks at the edible stimulus: 3.7 ± 1.0, at the non-edible stimulus: 1.6 ± 0.5).

Before the start of each trial we counted the number of guppies present in the selected group and randomly selected one individual as focal individual (because often, the fish would travel beyond the field of the camera, making problematic the recording of entire groups) of which we determined sex and size class (small/middle/large). Both stimuli were then slowly and simultaneously lowered into the water approximately equidistant from the focal individual (and whenever possible also equidistant from the whole group). Trials lasted 2 minutes following the introduction of the stimuli, giving the fish sufficient time to inspect the stimuli (although our main results did not change when considering shorter time periods (down to 30 s)). We scored the number of approaches and pecks made by the focal individual towards the edible and the non-edible stimuli. An approach was defined as the focal individual moving towards the stimulus within one body length from it and a peck was defined as the focal individual biting or nibbling the stimulus. A peck always followed either an approach or a previous peck whereas an approach was not necessarily followed by a peck. In total 607 trials were conducted, testing fish in group sizes ranging from 1 to 72 (mean group size = 6.2). Each trial was carried out in a different location along both streams and it is thus highly unlikely that the same groups and the same individuals were tested twice.

Analysis

We first investigated the effect of group size and predation level on sampling activity. For each randomly selected individual, we quantified every approach and peck toward each stimulus. To account for the high number of zeros in our data, we looked at the sampling activity using a hurdle model (count model with truncated negative binomial distribution and log link, zero hurdle model with binomial distribution and logit link) with group size, predation level, interaction between group size and predation level, river, size and sex as explanatory variables. We used likelihood ratio tests to compare the different models. We ran different models for approaches and pecks as they reflect qualitatively different decision processes despite both indicating a level of interest towards the stimuli.

We then studied decision accuracy, using the ratio (edible / (edible + non-edible)) as a measure of decision accuracy separating again approaches and pecks. We used this ratio as response variable in a generalized linear model with sampling activity (sum of approaches and pecks), group size, interaction between activity and group size, predation

level, river, size and sex as explanatory variables. We used binomial errors and a logit-link function since the response variable was bound between zero and one (with 0 indicating that all approaches/pecks were towards the non-edible stimulus, and 1 indicating that all approaches/pecks were towards the edible stimulus).

Finally, to minimize the effect of sampling activity and focus on the effect of group size, we also looked at the first decision made (i.e. first approach and first peck).

Results

Sampling activity

The number of approaches was not affected by group size (Fig. 3.1; count model: X^2 = 0.02, df = 1, p = 0.89). However, there was a significant effect of predation level (Fig. 3.1; count model: X^2 = 7.15, df = 1, p = 0.008; zero hurdle model: X^2 = 2.66, df = 1, p = 0.10), with fish in low predation areas approaching both types of stimuli more often than fish in high predation area. There was no significant interaction effect between group size and predation level (count model: X^2 = 0.22, df = 2, p = 0.90). There was also a significant difference between the two rivers (Fig. 3.1; count model: X^2 = 12.52, df = 1, p < 0.001; zero hurdle model: X^2 = 11.41, df = 1, p < 0.001), with fish in the Turure River being more likely to approach, and approaching more often than fish in the Aripo River. There was no effect of sex (count model: X^2 = 1.23, df = 1, p = 0.27) but there was an effect of the size of focal fish (Fig. 3.1; count model: X^2 = 4.31, df = 1, p = 0.032; zero hurdle model: X^2 = 7.27, df = 1, p = 0.007), with larger fish being more likely to approach both types of stimuli, and approaching them more often than smaller fish.

Group size did also not affect the number of pecks (Fig. 3.1; count model: X^2 = 1.08, df = 1, p = 0.30). There was, again, a significant effect of predation level (Fig. 3.1; count model: X^2 = 14.54, df = 1, p < 0.001; zero hurdle model: X^2 = 3.42, df = 1, p = 0.064), with fish in low predation areas pecking at both types of stimuli more often than fish in high predation areas. There was also no significant interaction effect between group size and predation level (count model: X^2 = 1.60, df = 2, p = 0.45). There was no significant difference between the two rivers (Fig. 3.1; count model: X^2 = 2.70, df = , p = 0.10). There was no effect of sex (count model: X^2 = 0.96, df = 1, p = 0.33) but there was an effect of the size of focal fish (Fig. 3.1; count model: X^2 = 5.01, df = 1, p = 0.025; zero

hurdle model: $X^2 = 29.13$, df = 1, $p < 0.001$), with larger fish being more likely to peck, and pecking more often than smaller fish.

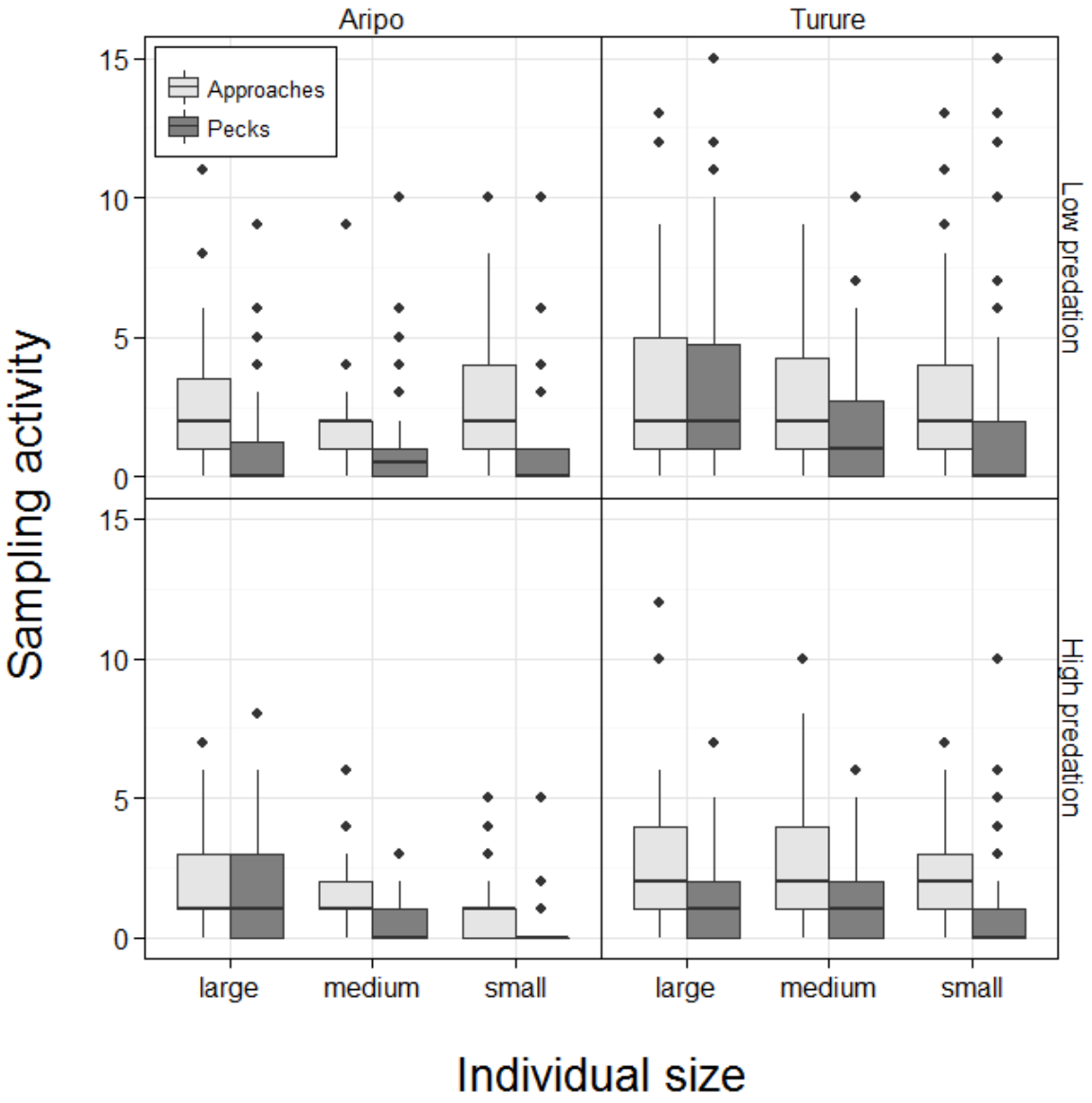

Figure 3.1: Sampling activity is greater in low predation sites (upper parts of the rivers) than in high predation sites (lower parts of rivers). It is also greater for larger individuals than smaller ones. Additionally, it is greater in the Turure River than in the Aripo River. Shown are medians and interquartile ranges. Data points outside 1.5 times the interquartile ranges are displayed separately.

Decision accuracy

The accuracy of approaches was affected both by sampling activity (Fig. 3.2a; glm: X^2 = 35.41, df = 1, $p < 0.001$) and by group size (Fig. 3.2b; X^2 = 16.61, df = 1, $p < 0.001$). Individuals which sampled more, and individuals in larger groups had a higher approach accuracy. There was no significant effect of interaction between sampling activity and group size on the accuracy of approaches (X^2 = 1.11, df = 1, p = 0.29). The accuracy of pecks was also affected both by sampling activity (Fig. 3.2a; glm: X^2 = 22.23, df = 1, $p <$ 0.001) and by group size (Fig. 3.2b; X^2 = 6.13, df = 1, p = 0.013), with individuals with a higher sampling activity and individuals in larger groups having a higher pecking accuracy. There was no significant effect of interaction between sampling activity and group size on the accuracy of approaches (X^2 = 0.08, df = 1, p = 0.77). Both for approaches and for pecks, there was no significant effect of predation level, river, size and sex. When using a different measure of decision accuracy, the so-called sensitivity index which combines the true and false positive rate of a decision maker and is widely used in signal detection theory (Macmillan and Creelman, 2004), we obtained quantitatively similar results.

First choice

Group size had a significant positive effect on the probability that the first approach is made towards the edible stimulus (glm: X^2 = 11.53, df = 1, $p < 0.001$). However, no effect of group size was found on the probability that the first peck is made towards the edible stimulus (glm: X^2 = 1.48, df = 1, p = 0.22).

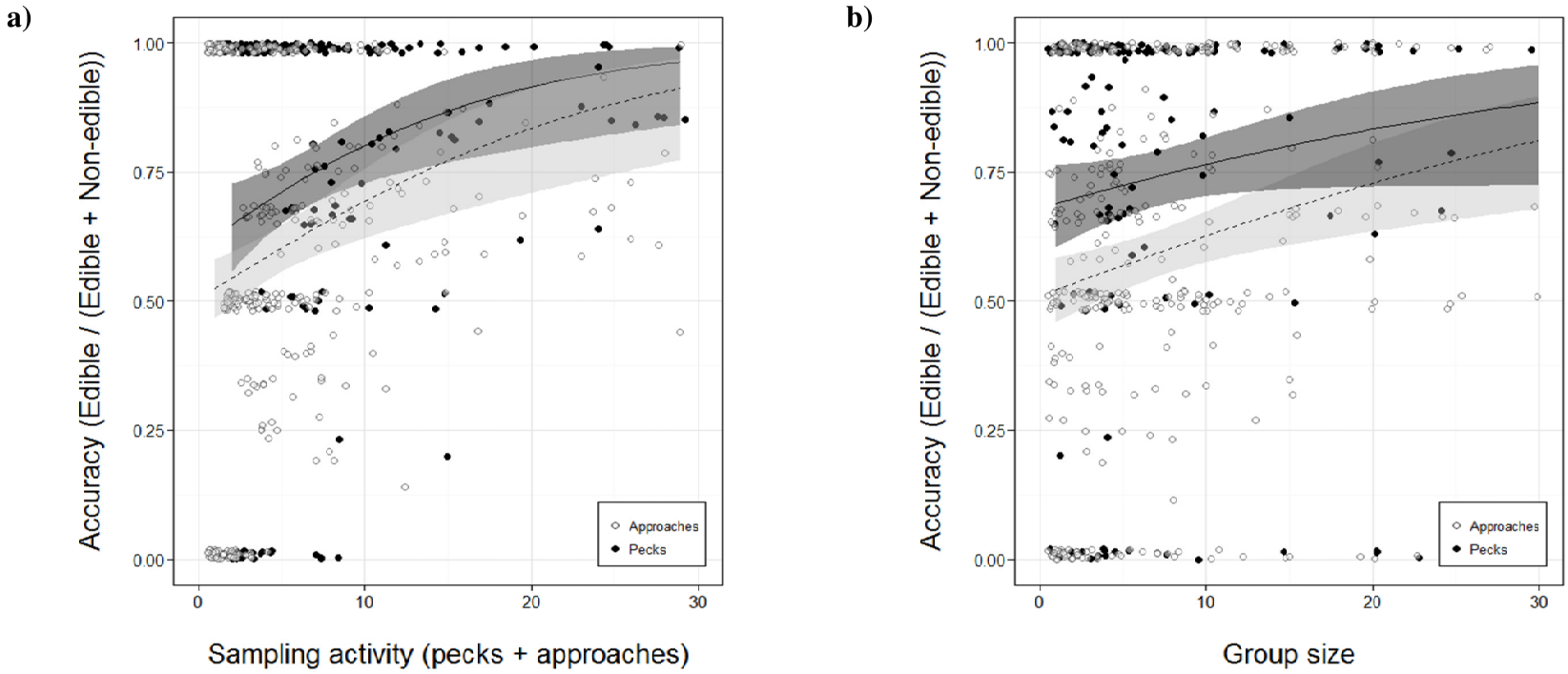

Figure 3.2: Accuracy increases as a function of **a)** sampling activity and **b)** group size. Approaches are represented in red and pecks are represented in blue. For clarity, 8 groups with group sizes larger than 30 are not shown but are included in the analysis and in the computation of the logistic regression, shown for pecks as a black line with 95 % confidence interval (dark shaded area) and for approaches as a dotted line with 95 % confidence interval (light shaded area).

Discussion

We found no correlation between sampling activity and group size. However, sampling activity was higher in low predation sites as predicted. Decision accuracy increased both with group size and sampling activity. Interestingly, we found no effect of predation level on decision accuracy, despite a higher sampling activity in low-predation sites.

Sampling activity was affected by several variables. It was greater in low predation habitats, which harbour fewer and smaller predator species (Magurran, 2005). This suggests that guppies were prone to spend more time and energy to individually sample potential food in low predation sites. However, we found no evidence of sampling activity being group-size-dependent in high-predation habitats; we had expected individuals to sample more when in bigger groups due to social facilitation and competition. Sampling activity was also greater for larger individuals, which are less at risk of being eaten by gape-limited predators (Croft et al., 2006; Mattingly and Butler, 1994). However, the effect of individual size on sampling activity did not translate into an effect on accuracy. Sex of the focal individual affected neither its sampling activity nor its accuracy, confirming that both sexes are attracted to carotenoids (Rodd et al., 2002) and not just exclusively males. Sampling activity was higher in the Turure River than in the Aripo River, which might be explained by ecological differences such as differences in the availability of carotenoids. This would be an interesting issue to explore in future studies.

Decision accuracy increased both with sampling activity and group size, confirming previous studies (e.g. (Brown and Laland, 2003; Brown et al., 2006; Liker and Bókony, 2009; Sumpter et al., 2008a; Warburton, 2003; Ward et al., 2008, 2012)). Surprisingly, predation level had no effect on decision accuracy, despite its effect on sampling activity. We can only speculate that under different predation regimes, private information and social information are used differently but lead to the same decision quality. Concerning the first decision made (before any private information had been gathered), the probability that the first approach was directed toward the edible stimulus increased with group size, suggesting that guppies used the behaviour of conspecifics in their choice of which stimulus to approach first. This result is consistent with previous work showing that ninespine sticklebacks (*Pungitius pungitius*) rely on the presence and feeding rate of their conspecifics to decide on which food patch to approach (Coolen et al., 2005; Webster and Laland, 2013), and with studies on gaze following which indicate

that social information strongly guides individual attention in humans (Gallup et al., 2012). However, the first peck was not influenced by group size, which suggests a different underlying mechanism for this decision.

In our study, individuals experienced two stimuli at the same time and made their decision over a time period that allowed individuals to sample both stimuli, contrary to other studies on group decision-making that usually involve the detection of a single stimulus (e.g. a predator (Ward et al., 2008) or a shaded area (Berdahl et al., 2013)) within a short period of a few seconds which is too brief for individuals to explore all available options (but see Miller et al., 2013). In our study, sampling activity was an important part of the decision process and this enabled us to quantify this factor as a proxy for private information that is available to individuals. The use of social information was inferred from a positive correlation between decision accuracy and group size in the absence of any increase in sampling activity (private information) with group size, and from our data showing a group size trend for first-choice approach (towards the edible stimulus).

Previous studies have shown that individuals in groups can outperform single individuals at cognitive tasks, such as detecting and avoiding predators and detecting food patches (Cresswell, 1994; Kenward, 1978; Krause and Ruxton, 2002; Lima, 1994, 1995a; Ward et al., 2008, 2011). The present study is novel as it shows collective cognition in multiple populations that are subject to different ecological conditions. We found that both private and social information contribute to decision accuracy but we were not able to separate out their relative contribution which might be an exciting field for future studies (see Arganda et al. (2012) for a lab study). Our results on first approach and first peck do show, however, that the use of private and social information appears to be context-dependent.

Ethics statement: This research was performed in accordance with the laws, guidelines and ethical standards of the country in which they were performed (Trinidad).

4. Information transmission via movement behaviour improves decision accuracy in human groups

Clément RJG, Wolf M, Snijders L, Krause J & Kurvers RHJM

A major advantage of group living is increased decision accuracy. In animal groups information is often transmitted via movement. For example, an individual quickly moving away from its group may indicate approaching predators. However, individuals also make mistakes which can initiate information cascades. How responsive should individuals then be to escaping group members? Increasing responsiveness increases true positives (i.e. escape when a predator is present) but at the cost of increased false positives (i.e. escape when a predator is absent). Conversely, reducing responsiveness decreases not only false positives but also true positives, resulting in a fundamental trade-off in decision accuracy. Here we investigated how socially responsive individuals are to information transmission via movement: a gradual mechanism of information transmission abundant in nature. We performed a simulated predator detection task using human groups in which humans stepped forward if they wanted to escape. We confirm that this simple movement mechanism allows individuals in groups to simultaneously increase true positives and decrease false positives. The increase in the number of escapees over time during collective decisions depended on the personal information of the group members: Individual predator detection by only a few group members rarely resulted in anyone stepping forward. Individual predator detection by a quarter of the group often resulted in the entire group escaping. Finally, individual predator detection by at least half of the group led to a rapid escape of the whole group. Overall, the increase in the number of escapees over time followed a linear response. Since information transmission via movement is widespread in animal groups, this mechanism is expected to be relevant for many animal groups to improve decision accuracy.

Animal Behaviour 105: 85-93, 2015

Introduction

Group living has evolved owing to the many advantages it provides to the individuals that are part of a collective, such as increased safety and increased opportunities for detecting food or finding mates (Krause and Ruxton, 2002; Sumpter, 2010). Such advantages often result from the ability of groups to achieve higher decision accuracy than single individuals (Clément et al., 2013; Conradt and List, 2009; Ward et al., 2011).

Animal groups frequently need to make consensus decisions in order to maintain group cohesion and its associated benefits and, in many cases, the information exchange underlying these decisions takes place via movement (Conradt and Roper, 2003; Miller et al., 2013; Strandburg-Peshkin et al., 2013). For example, in fish shoals and bird flocks, an individual moving away from the group indicates to the others its intention to change direction or leave a current location (Beauchamp and Ruxton, 2007; Radakov, 1973). Changes in speed and direction are often the primary ways of information transmission in large groups (Handegard et al., 2012; Katz et al., 2011; Sumpter et al., 2008b). Although verbal communication often plays a crucial role in humans, movement alone can also serve as the sole cue in everyday human interactions, as observed in pedestrians at road crossings, where individuals are more likely to jaywalk after seeing another individual doing so (Faria et al., 2010a).

Individuals moving in a given direction can thus provide information to their group members, which can increase collective accuracy (Berdahl et al., 2013; Lazarus, 1979; Treherne and Foster, 1981; Ward et al., 2011). In situations involving predation risk, for example, a sudden movement away from the group by some individuals usually indicates the presence of a predator, as shown in fish schools and bird flocks, and can trigger the reaction of the whole group (Cresswell, 1994; Hingee and Magrath, 2009; Kenward, 1978; Lima, 1994; Radakov, 1973). However, individuals might also make mistakes (e.g. false alarms) and this can give rise to cascades of false information, whereby the whole group is led into a wrong action (Bikhchandani et al., 1992; Conradt, 2011; Giraldeau et al., 2002; Janis, 1982). This raises the question of how sensitive should individuals be to information from their conspecifics, a classic dilemma for decision makers under uncertainty (Beauchamp and Ruxton, 2007; Giraldeau et al., 2002; Lima, 1995b; McNeil et al., 1975; Swets, 1988; Swets et al., 2000; Zweig and Campbell, 1993). The aim of a decision maker is to take an action whenever a particular condition in its environment is fulfilled, but abstain from it when the condition is not fulfilled. In the

case of predation risk, for example, an animal should run away in the presence of a predator but not in its absence. However, cues indicating a particular condition (e.g. presence of a predator) may also appear in its absence. Conversely, the condition may be fulfilled without any obvious cues. Increasing responsiveness to such cues leads decision makers to increase their chances of correctly taking the action when the condition is fulfilled (making a true positive, e.g. run away in the presence of predators) but also that of erroneously taking this action in its absence (committing a false positive, e.g. run away in the absence of predators) while reducing responsiveness leads to a decrease in false positives, but at the cost of reduced true positives. This fundamental trade-off in decision accuracy under uncertainty is encountered across many different contexts, including predator detection (Beauchamp and Ruxton, 2007; Lima, 1995b) and food detection (Giraldeau et al., 2002).

Recent research has shown that individuals in groups can overcome this fundamental trade-off. Using a mathematical model, Wolf et al. (2013) predicted that, compared to solitary individuals, individuals in groups can simultaneously increase true positives and decrease false positives whenever individuals employ a quorum threshold in between the average true and false positive rates of the other group members. This prediction was then tested with groups of humans in a simulated predator detection experiment in which participants anonymously expressed their intention to either stay or escape using voting machines, after which they received a summary chart showing the aggregated decisions of all group members and could decide again. The experiment revealed that individuals indeed used a quorum threshold in between the average true and false positive rates of the other group members, thereby simultaneously increasing true positives and decreasing false positives.

However, in many animal groups an individual does not have access to one aggregated response consisting of all the combined independent decisions, but information becomes gradually available. An example of this can be found in antipredator behaviour, in which often one or a few individuals make a decision (i.e. escape), upon which others can decide to either follow this decision or not. Also, in many animal groups, individuals can decide to readjust their decision, if they notice that their decision is not followed by others. Moreover, in most animal groups movement is the prime cue of information transfer. Therefore, we developed an experiment that resembles a more realistic scenario which could be relevant to many social animals. We performed a predator detection experiment in which individuals moved spatially to indicate their

choice, allowing the information to be transmitted much more gradually and dynamically. We hypothesized that a group of individuals that can only use movement to indicate preferences is able to increase true positives and decrease false positives. Moreover, we predicted that groups in which only a few individuals detect a predator individually (i.e. low true positive rate) would need a lower fraction of the group indicating escape to cause the group to escape compared to groups in which a large number of individuals detected the predator individually (i.e. high true positive rate). This was predicted because in groups with few individuals detecting a predator, we expected individuals to learn that even low numbers of people escaping can correctly indicate the presence of a predator.

Methods

Experimental set-up

Students were recruited from the University of Bielefeld (Germany) and Wageningen University (The Netherlands). The 310 participants in the tests were distributed over 15 groups (average group size 20.7, range 17–23). Informed consent was obtained from all participants prior to the experiment and data collection was anonymous.

Each group was confronted with the following predator detection experiment, resembling the experiment described in Wolf et al. (2013). All individuals in a group were instructed to stand behind a line (4.5 m away from the screen, Fig. 4.1a) and for 2 s a slide showing 144 fish (aligned in a 9 x 16 grid) was projected onto a white screen (Fig. 4.1b). All fish in this school were identical, except one odd fish, which had either six or seven spines (Fig. 4.1c). The other 143 fish had no spines. The seven-spined fish represented a dangerous predator, whereas the six-spined fish represented a harmless individual, akin to a natural situation in which individuals have to discriminate between harmful and harmless shapes (see e.g. Beauchamp, 2010; Cresswell et al., 2000). We instructed our subjects to adopt the following decision rule: “If you see no odd fish or an odd fish with six spines then it is safe and you should stay. If you see an odd fish with seven spines then it is dangerous and you should escape”. Participants saw the slide of the fish school only once, for 2 s, after which they had 5 s to take an individual decision (polling 1) using an electronic keypad (Key Point Interactive Audience Software for Power Point, version 2.0.142 Standard Edition), ensuring independent votes of participants. Individuals were asked to press 1 if they wanted to escape and they were not

allowed to move, gesture or communicate during this stage. Individuals did not receive information about the results of polling 1.

After polling 1, the participants were allowed to make a second decision (polling 2). Two parallel lines, 1 m apart, had been drawn on the floor (Fig. 4.1a). At the beginning of each trial, participants stood behind the line furthest away from the screen. Individuals were asked to stay behind the line if they wanted to stay, or to step forward (crossing the two lines drawn on the floor) if they wanted to escape (see Fig. 4.1a). Participants had 12 s to make a final decision and were allowed to move back and forth as often as they wanted during this time. Individuals were not allowed to communicate or gesture during polling 2, but they were able to observe and influence each other by taking into account the movement of their group members. After 12 s, we counted the participants that decided to stay and escape. We also recorded the movement behaviour using a video camera (Sony HDR-XR520V, 25 frames/s) mounted on an elevated tripod next to the screen facing the volunteers in order to get the widest angle and avoid some of the volunteers' movement being masked by others (Fig. 4.1a). After polling 2, we presented the correct answer on the screen and instructed all participants to move back behind the original line upon which a new round would start. To motivate participants, we had a small reward (chocolate bars) for the participants of the group with the highest proportion of correct decisions. This was communicated to the participants before the start of the experiment.

There were two treatments: one fish with six spines and 143 fish without spines or one fish with seven spines and 143 fish without spines. Each treatment was replicated 14 times resulting in a total of 28 rounds per group. The treatment order and the position of the odd fish were randomized. Prior to the 28 rounds we performed two test rounds during which we instructed the participants about the procedures. The results of the two test rounds were not included in our analysis. Both the six- and the seven-spined fish came in six different forms (spines arranged differently over the body of the fish) to avoid habituation to a specific search image. Each form appeared two to three times throughout the 28 rounds.

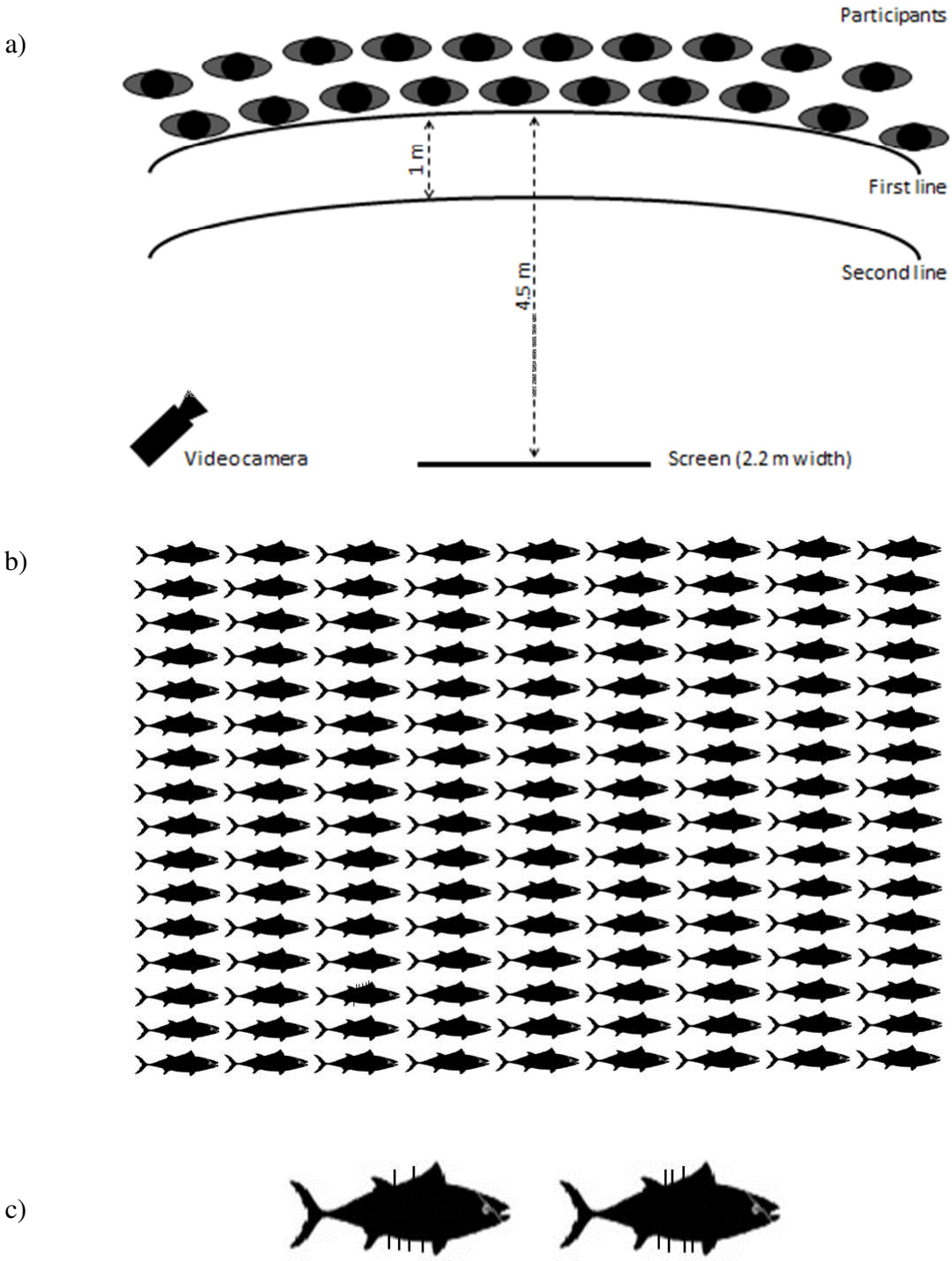

Figure 4.1: (a) Experimental arena. All participants started behind the first line and observed for 2 s (b) a group of 144 fish containing either (c) a six-spined ("nondangerous") or seven-spined ("dangerous") fish projected on the screen. See Methods for details of the procedure.

Analysis

We analysed our data in three steps. First, we studied whether social information improved decision accuracy. We compared the average true and false positive rates per group for polling 1 and 2 using a Wilcoxon signed-rank test in R version 3.1.1 (R Development Core Team, 2015). Note that a true positive is achieved whenever a seven-spined fish is present and a participant decides to escape; a false positive is committed whenever a six-spined fish is present and a participant decides to escape. We also evaluated whether individuals became better in correctly detecting the predator (individual detection) and/or better in using the social information. For each group we calculated general accuracy per round and per polling. General accuracy was calculated as the fraction of correct decisions over the total number of decisions and is thus a combination of true positives and true negatives. General accuracy was used as the response variable in a generalized linear mixed model with group (categorical) as a random term; the fixed effects were round number (continuous), polling (categorical: individual versus collective), treatment (categorical: predator present or absent) and the interaction between round number and polling. In the model we used binomial errors and a logit-link function since the response variable followed a distribution between 0 and 1.

Second, we studied the dynamics of the decision-making process during polling 2. From the recordings, using JWatcher 1.0 (Blumstein et al., 2006), we scored per round and per group the time point at which each individual crossed the second line ("escaped") and also whether individuals stepped back again behind the first line (i.e. changed their mind from "escaping" to "staying"). From these data, we calculated for each round the "decision time", defined as the time point after which no individual changed its decision. We then calculated for each group and each round the slope of the linear regression of the cumulative escapes as a function of time, which gives an estimation of the speed of escape (expressed in number of people escaping/s). To compute the speed of escaping behaviour, we only focused on trials in which at least five people escaped during polling 2 since trials in which up to four people escaped did not present a real building-up of number of escapees over time and a linear regression computed from values of four or less would be meaningless. Moreover, the design of the experiment did not allow an estimation of the speed of the decision when the decision was staying, as only escaping required an active decision and was measurable. Owing to video failure on one of the two

experimental locations, this part of the analyses is limited to seven groups (150 individuals).

Third, we compared the relationship between the fraction of individuals indicating escape in polling 1 and the fraction of individuals escaping at the end of polling 2 between groups. Although individuals did not receive information about the outcome of polling 1, we could still study whether groups differed in how much influence a certain mass had in mobilizing the group. We investigated whether the relationship between individuals escaping in polling 1 and individuals escaping in polling 2 followed a quorum-like pattern (i.e. a sigmoidal curve). A quorum response implies that the probability that an individual decides to escape during polling 2 increases in a step-like way with the number of other individuals that decided to escape in polling 1. We formalized this relationship for each group using the following equation (Conradt, 2012; Sumpter and Pratt, 2009):

$$p = \frac{x^k}{x^k + q^k}$$

where p is the probability that an individual's final decision during polling 2 is to escape, x is the fraction of individuals indicating escape during polling 1, q is the threshold value at which the response has the steepest increase and k determines the steepness of this increase. As a rule of thumb, a quorum response (corresponding to a strong sigmoidal curve) occurs if $k \geq 2$ and higher values of k imply stronger quorum responses (Conradt, 2012; Sumpter and Pratt, 2009). In all groups the k value was higher than 2 with an average value of 12.7 (median 5.17, range 3.03–109.6). We then estimated the q value for each group and compared the q values to the true and false positive rates of polling 1 using Spearman's rank correlation coefficient.

Results

The true positive percentage was higher in polling 2, where social information was available, than in polling 1, where individuals decided independently (mean ± SE: polling 1: 41.0 ± 8.6, range 23.5–53.6; polling 2: 70.7 ± 11.1, range 51.7–89.6; $Z = 3.408$, $P < 0.001$). All 15 groups had higher true positives in polling 2 than in polling 1 (Fig. 4.2a). The false positive percentage was lower in polling 2 than in polling 1 (polling 1: 9.2 ± 2.6, range 2.7–12.9; polling 2: 4.6 ± 3.6, range 0.7–13.7; $Z = -3.237$, $P < 0.001$). Of the 15 groups, 13 had a lower false positive percentage in polling 2 than in polling 1 (Fig.

4.2b). There was no significant effect of round (estimate ± SE = -0.018 ± 0.020; z = -0.932, P = 0.351) or of the interaction between round and polling (estimate ± SE = -0.009 ± 0.027; z = -0.345, P = 0.730) on general accuracy suggesting that individuals did not become gradually better at correctly detecting the presence/absence of a predator (i.e. individual detection) or at using the social information. There was an effect of polling (estimate ± SE = -1.542 ± 0.470; z = -3.287, P = 0.001) with individuals achieving higher accuracy in polling 2 than in polling 1. There was also an effect of treatment (estimate ± SE = -4.653 ± 0.472; z = -9.853, P < 0.001): individuals were better at staying when no predator was present than at escaping when a predator was present (see also Fig. 4.2a, and 4.2b).

Fig. 4.3 shows the increase in the number of escapees over time per round for the seven groups. For clarity, the panels are ranked (within groups) according to the number of individuals indicating escape during polling 1. Moving from the left to the right side of the figure, we note that with no, or very few, individuals indicating escape in polling 1, nobody stepped forward during polling 2. With an increasing number of individuals deciding to escape during polling 1, there was some building up of low numbers of individuals, but this generally did not lead to much recruitment and occasionally participants changed their mind and stepped back again (although this only happened 18 times over all 196 trials; see for examples panels 7 and 12 of group 4). A further increase in the number of people indicating escape in polling 1 resulted in an increase in the number of escapees during polling 2, sometimes leading to the recruitment of the entire group. Finally, on the right side of the figure, a large number of people indicating escape during polling 1 resulted in the entire group moving forward very fast.

We then studied how the fraction of individuals indicating escape in polling 1 affected (1) the fraction of individuals escaping in polling 2, (2) the "decision time" and (3) the number of people escaping/s (only including trials in which at least five people escaped, N = 81 trials). The relationship between the fraction of individuals escaping during polling 1 and the fraction of individuals escaping at the end of polling 2 followed a sigmoidal pattern (Fig. 4.4a, 4.5a). When looking at the decision time in more detail (Fig. 4.4a), we note that a very low (< 0.1) and a high (> 0.5) fraction of individuals indicating escape in polling 1 gave rise to rapid decisions (purple shading), whereas a moderate fraction of individuals indicating escape in polling 1 (0.2–0.5) led to long decision times (blue shading). The increase in the number of escapees over time within each round (i.e. Fig. 4.3) generally followed a linear pattern (average R^2 linear regression: 0.95 ± 0.08,

range 0.54–1.00). Note that the linear fit was not perfect in all cases, and some cases could be better described with more complex functions. For simplicity, we restricted our analysis to linear relationships which captured most of the dynamics fairly well. We calculated the slope of this relationship per round per group (only including trials in which five or more participants escaped). The slope depended on the fraction of individuals deciding to escape during polling 1: the higher this fraction the steeper the slope (estimate ± SE = 7.68 ± 0.74; $t = 10.41$, $P < 0.001$; Fig. 4.4b).

Finally, we looked at differences between groups. We compared the relationship between the fraction of individuals escaping during polling 1 and 2 between groups (Fig. 4.5a). There was a positive correlation between the average true positive rate of the group during polling 1 and the q value (threshold value at which the response has the steepest increase; $r_S = 0.559$, $P = 0.030$; Fig. 4.5b), suggesting that individuals adjusted their sensitivity adaptively to the performance of the group. There was no correlation between the average false positive rate of the group during polling 1 and the q value ($r_S = 0.446$, $P = 0.095$), probably because of the lower between-group variation in false positives than true positives during polling 1.

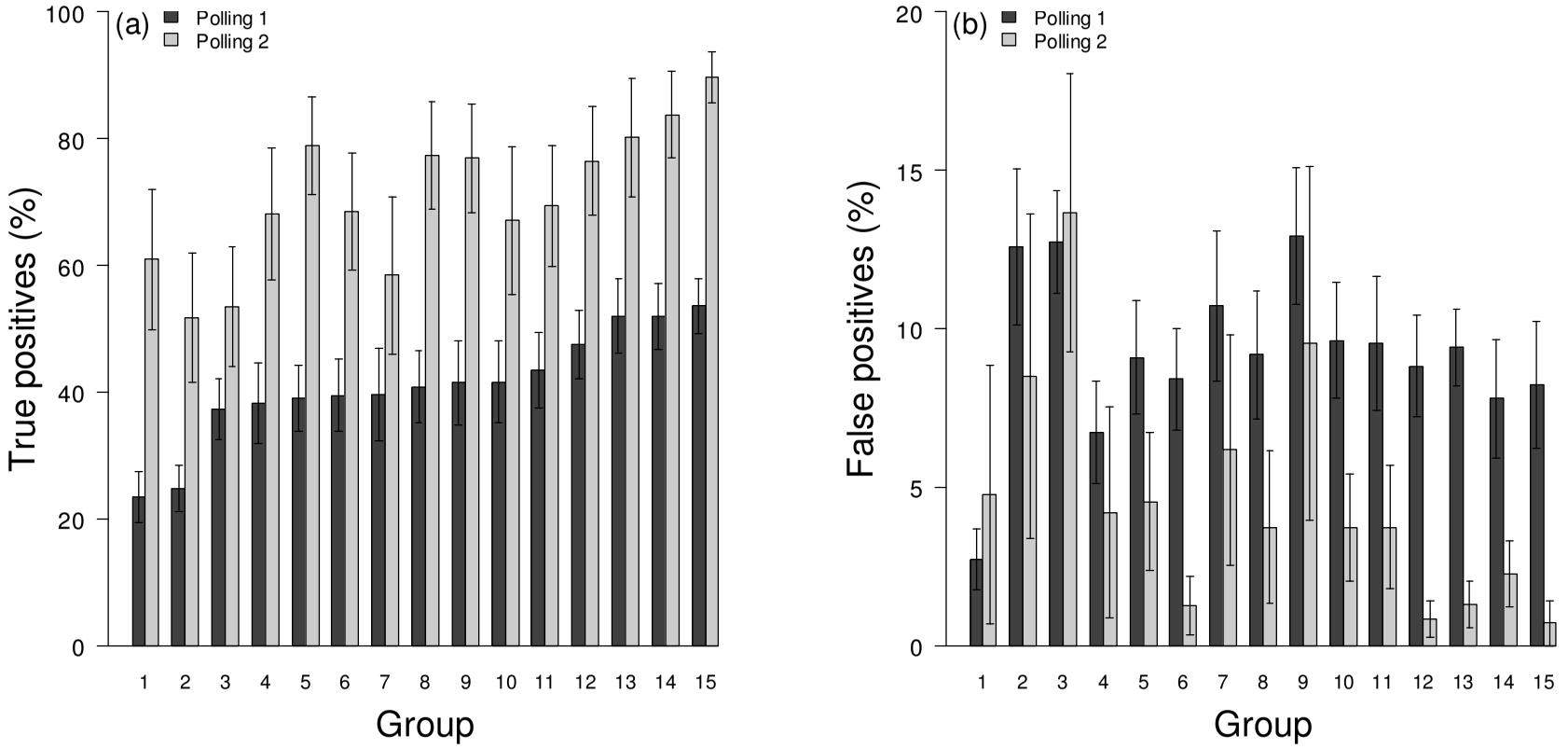

Figure 4.2: The mean percentage ± SE of (a) true positives and (b) false positives of individuals before (polling 1) and after (polling2) taking into account the decision of other individuals. For clarity, the groups in (a) and (b) are ranked according to the true positive score in polling 1.

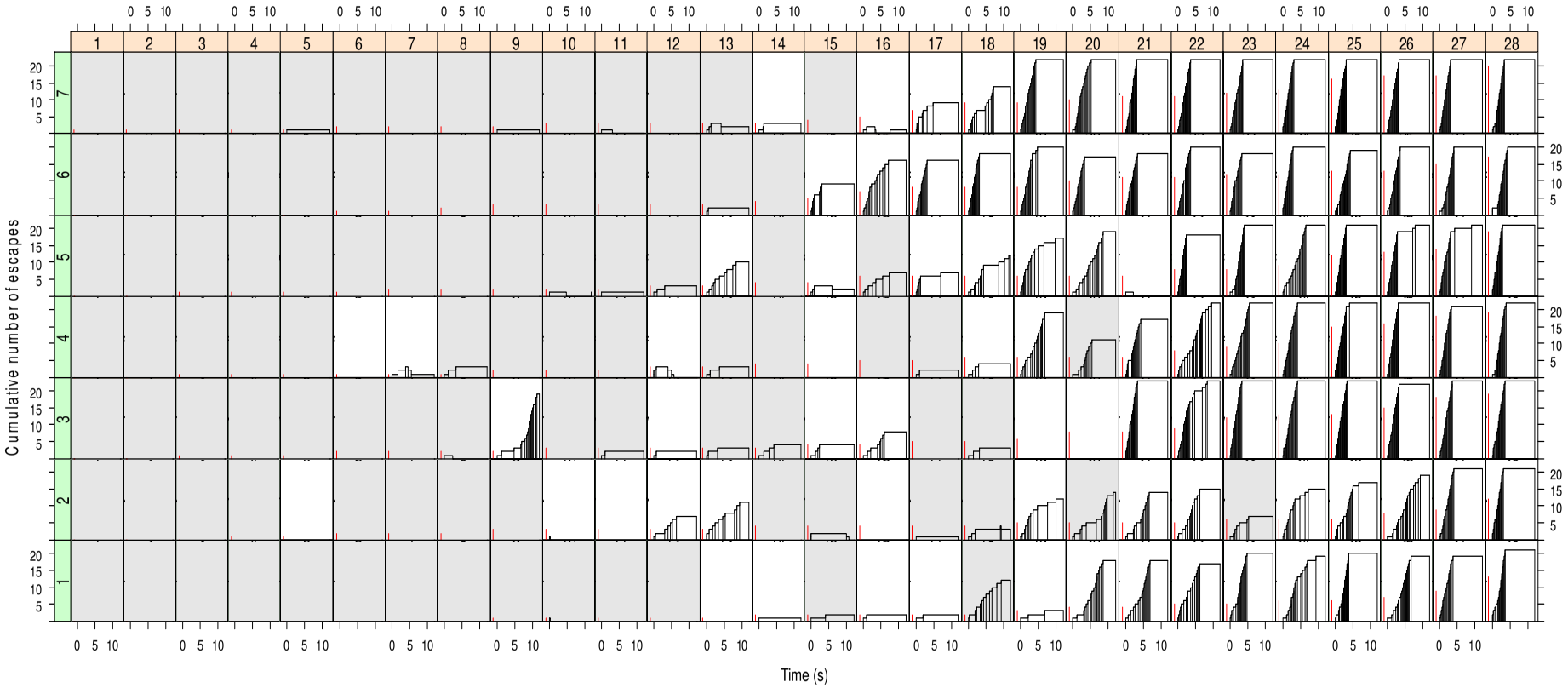

Figure 4.3: The number of escapees in relation to time during polling 2 per round for seven groups. Each panel shows the number of people escaping (y-axis) over the 12 s (x-axis). The bar on the left of each panel indicates the number of individuals indicating escape during polling 1. For clarity, the rounds are ranked according to the number of individuals indicating escape during polling 1. White panels indicate "predator present" (i.e. true positives) and grey panels indicate "predator absent" (i.e. false positives).

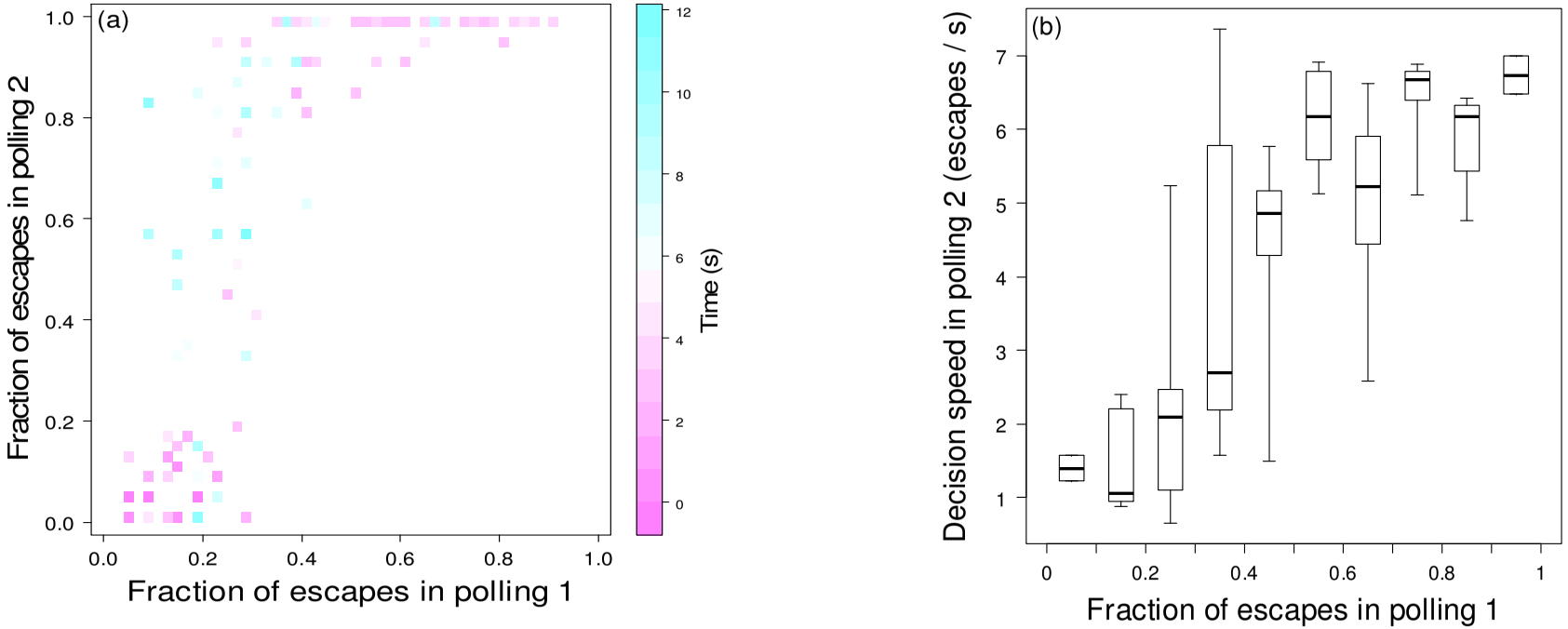

Figure 4.4: (a) The relationship between the fraction of individuals indicating escape during polling 1 and the fraction of individuals escaping at the end of polling 2. The colour of the square represents the decision time (i.e. the time after which no individual changed its decision) with blue colours representing long decision times and purple colours short decision times. Note the resolution of 0.02 per square for clarity, meaning that the colour of some of the squares is the result of averaging several trials that had very close values for polling 1 and polling 2. However, this did not affect the general pattern as the values for time to escape were also very close. (b) The relationship between the fraction of individuals indicating escape during polling 1 and the average number of participants escaping/s. Box plots show median values, 25th and 75th percentiles (boxes) and minimum and maximum values (whiskers). For (b) we only included trials in which at least five

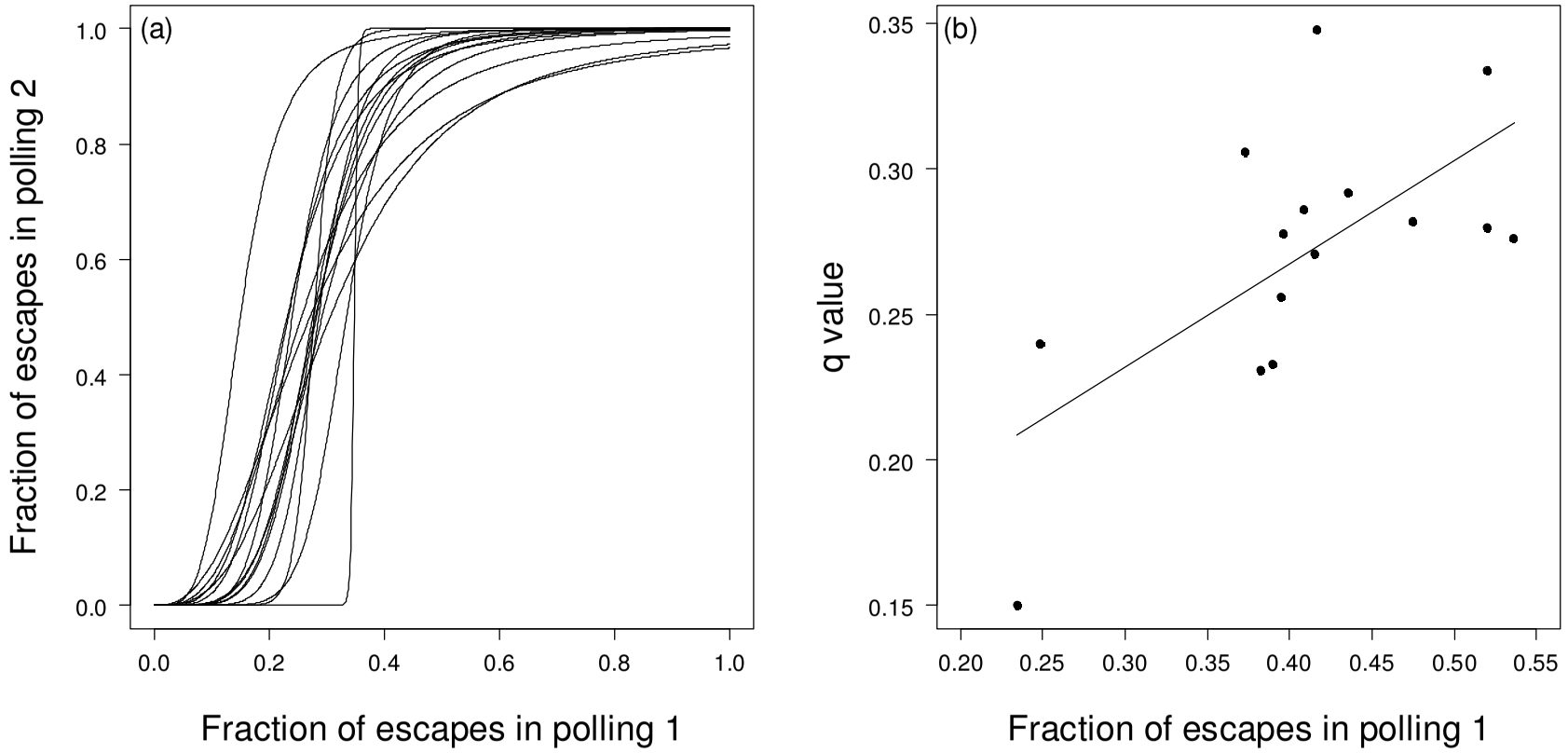

Figure 4.5: (a) The relationship between the fraction of individuals indicating escape during polling 1 and the fraction of individuals deciding to escape at the end of polling 2 per group (N = 15 groups) using the estimated q and k values of each group. The q value of each group corresponds to the x-value at which the curve has the steepest increase (which happens at a y-value of 0.5). (b) The relationship between the q value and the average true positive rate during polling 1.

Discussion

Our results show that individuals in groups can simultaneously increase true positives and decrease false positives. This result is similar to the findings of Wolf et al. (2013) but here we show that this can arise by using a simple and generic movement rule and a gradual building up of information. This is a more realistic scenario for most animal groups. Information transmission via movement alone is a very common mechanism in gregarious animals observed in deer and primates (Conradt and Roper, 2003; Sueur et al., 2010), as well as bird flocks (Beauchamp and Ruxton, 2007; Cresswell, 1994; Lima, 1994), fish schools (Herbert-Read et al., 2011; Radakov, 1973; Ward et al., 2011) and even human pedestrians (Faria et al., 2010a). Using a movement scenario thus allowed individuals to express their opinions to others via a simple and generic decision mechanism.

When looking in detail at the decision-making process, we see that individuals rarely stepped forward alone. When there were few individuals indicating escape during polling 1, they generally did not step forward by themselves. Individuals thus did seem to be socially inhibited about disclosing their personal information which highlights the important distinction between the opinion of an individual and its actions in a social context (Weber, 1978). Most likely, individuals were monitoring whether others were willing to step forward and only moved forward if enough others were willing to join this decision. At the other extreme, if half, or more, of the group members indicated escape during polling 1, then the decision-making process during polling 2 became very predictable and all participants stepped forward quickly. In between, the process was less predictable, and here other important factors might have played a role such as self-confidence of individuals and/or certainty of information, making the decision process longer. Interestingly, the number of escapees over time during the collective decision-making process generally followed a linear pattern (i.e. Fig. 4.3). However, if we only take into account the results of polling 1 (individual decision) and of polling 2 (the collective decision), we see that a linear building up can still lead to a quorum response (see Fig. 4.5a). Without the personal information of the individuals we would have thus concluded that there is no quorum response, but quorum responses do apparently play a role. It is important to note that here we focused on individuals crossing the second line as a potential cue for others, but that individuals might use a variety of movement cues as social information. Individuals might, for example, use stepping forward by others as a

cue, the speed of stepping forward and/or a combination of different cues. Although beyond the scope of this study, it would be an interesting avenue for further research to look in more detail at which (combination of) movement cues are used.

It is well known that groups of decision makers employing a majority vote can outperform solitary decision makers whenever the individual probability of making a correct decision is higher than 50%. This is known as the Condorcet theorem (Boland, 1989; List, 2004). A corollary of this theorem is that, if the individual probability of making a correct decision is lower than 50%, increasing the number of decision makers will decrease the accuracy of the group decision. Interestingly, the average true positive rate of individual decisions in our experiment was maximally 53.6% and could be as low as 23.5% (see Fig. 4.2a). Nevertheless, in all collective decisions true positive rates went up dramatically (see Fig. 4.2a) showing that our participants did not merely follow a simple majority rule. This also became apparent when calculating the q values. These values were all well below 50% (range 15–34.8%) indicating that minorities were frequently followed. Our results also strongly suggest that individuals adjusted their thresholds to the average individual true positive rate of the group they were in. In groups with a very low average individual true positive rate, relatively few individuals indicating escape in polling 1 were required to mobilize group members to escape in polling 2. In contrast, in groups with a high average individual true positive rate, many individuals indicating escape in polling 1 were required to mobilize the others (Fig. 4.5). The result that small informed minorities were able to mobilize a large part of the group also confirms previous models (Couzin et al., 2005) and studies (Dyer et al., 2008, 2009; Faria et al., 2009; Miller et al., 2013) that showed that a small minority of informed leaders is able to lead a whole group in the right direction.

By enabling dynamic interactions between participants, our experimental set-up allowed for more realistic mechanisms of information exchange between the individuals involved in collective decision-making than in previous work. It also gave group members the opportunity to be influenced by highly dominant individuals or top performers (possessing private information of higher quality due to, for example, better eyesight; King, Cheng, Starke, & Myatt, 2011; King, Johnson, & Van Vugt, 2009). An interesting issue for further study would be to identify the precise mechanism allowing group coordination. Are there local effects, whereby neighbours of fast-moving individuals get dragged along, or are there only global effects? Is it a pure voting procedure as discussed in Wolf et al. (2013)? Is speed related to accuracy? Indeed, several

mechanisms could produce the result we observed. For example, especially in a predator avoidance context, individuals might differ in the timing of their decisions depending on their information status (Hilton et al., 1999; King and Cowlishaw, 2009). Also, confidence levels of individuals may vary across the different rounds, leading to different individuals alternating in taking the lead from one round to the other. In addition, all members had the same interest in escaping when there was a predator, and timing did not affect the payoff. As suggested in Kurvers et al. (2014a), an interesting investigation would be to include competition by granting higher rewards to the faster correct decision makers compared to the slower ones, as an early escape is clearly beneficial in the wild (Bednekoff and Lima, 1998). Finally, how do the dynamics of the decision-making process depend on group size and on the costs of specific errors (e.g. for animals in the wild, the costs of a false negative are usually higher than the costs of a false positive).

Next to future experiments in the human domain, we also envisage researchers studying the combination of true and false positives in nonhuman animal groups. As yet, these studies are largely absent. In the collective predator detection literature, for example, most studies have investigated either true positives (Lazarus, 1979; van Schaik et al., 1983; Sirot, 2006; Treherne and Foster, 1981; Ward et al., 2008, 2011) or false positives (Beauchamp, 2010; Beauchamp and Ruxton, 2007) but rarely the combination of true and false positives (Proctor et al., 2001; Roberts, 1997). Therefore it is currently unclear whether this decision mechanism (i.e. increasing true positives while decreasing false positives) is also operating in nonhuman animals. Given the inherent trade-off between true and false positives, studying one aspect will only provide part of the picture and to understand the complete costs and benefits of collective predator detection it is important to study both aspects simultaneously. Technological advancements such as accelerometers combined with GPS (Eagle and Pentland, 2006) that can capture the movement behaviour of multiple individuals simultaneously and can provide detailed information on how individuals react to decisions and movements of groupmates can prove useful in exploring these issues. An important difference to point out for many animal groups under predation is the potential benefit of staying together even in situations in which the group makes wrong decisions. Even when a predator attacks, it is usually more beneficial to stay in the group rather than leave it with the intention of escaping from the predator (Krause and Ruxton, 2002). This is an issue that we have not included in our experimental setting but is most likely to be an important mechanism in animal groups under predation.

Finally, we found that individuals were better at correctly staying than at correctly escaping. This could be because individuals had to make an active decision to escape (press a button in polling 1 and move forward in polling 2) but no active decision was required to stay. We decided to use this set-up because we wanted to create a situation comparable to animal groups in which the default behaviour is to stay (and continue foraging) but an active decision is required to escape.

Our study shows that information transmission via movement behaviour allows individuals in groups to simultaneously increase true positives and decrease false positives, thereby escaping an important trade-off in decision-making. Since information transmission via movement is widespread in animal decision-making, this mechanism could be relevant for a wide range of animal groups to improve decision accuracy. Our study also shows that there can be a discrepancy between the opinion of a person and his or her actions, which calls for more research in the human domain about the link between both. It is also a call for researchers studying animal decision-making to extract more information about the internal state of an individual and not only observe the decision outcome.

5. Collective cognition in humans: Groups outperform their best members in a sentence reconstruction task

Clément RJG, Krause S, Engelhardt N, Faria JJ, Krause J & Kurvers RHJM

Abstract

Group-living is widespread among animals and one of the major advantages of group-living is the ability of groups to solve cognitive problems that exceed individual ability. Humans also make use of collective cognition and have simultaneously developed a highly complex language to exchange information. Here we investigated collective cognition of human groups regarding language use in a realistic situation. Individuals listened to a public announcement and had to reconstruct the sentence alone or in groups. This situation is often encountered by humans, for instance at train stations or airports. Using recent developments in machine speech recognition, we analysed how well individuals and groups reconstructed the sentences from a syntactic (i.e., the number of errors) and semantic (i.e., the quality of the retrieved information) perspective. We show that groups perform better both on a syntactic and semantic level than even their best members. Groups made fewer errors and were able to retrieve more information when reconstructing the sentences, outcompeting even their best group members. Our study takes collective cognition studies to the more complex level of language use in humans.

Plos One 8: e77943, 2013

Introduction

Group-living is widespread among animals and one of the major advantages of group-living is the ability of groups to solve cognitive problems that exceed individual ability (Bonabeau et al., 1999; Conradt and List, 2009; Couzin, 2009; Krause and Ruxton, 2002; Krause et al., 2010; Wolf et al., 2013). This process is known as the many wrongs principle (Simons, 2004), swarm intelligence (Bonabeau et al., 1999; Krause et al., 2010), wisdom of crowds (Surowiecki, 2004) or collective cognition (CC) (Couzin, 2009). Fish, for example, make faster and more accurate decisions in groups than when alone (Ward et al., 2011), in ants larger colonies are faster at finding the best nesting sites (Franks et al., 2006) and in birds larger groups are more successful in innovative problem solving (Liker and Bókony, 2009). Also humans can make use of CC and CC has been shown to solve a number of different problems including predicting the results of elections (Arrow et al., 2008; Wolfers and Zitzewitz, 2004) solving letters-to-numbers problems (Laughlin et al., 2003, 2006) and increasing speed and accuracy at reaching a target when navigating as a group (Faria et al., 2009).

A remarkable feature of humans is the use of a highly complex language. Language is thought to have played a critical role in the evolution of hominids (Noble and Davidson, 1996) giving them a unique way of sharing information among conspecifics. Moreover, group discussion is still the most widely used method by human groups to arrive at consensus decisions. Several studies have investigated CC of human groups with regards to quantity estimations and letters-to number problems (see Krause et al. (2010) for a review). However, few studies made language itself the focus of their investigation. Here we simulated a realistic scenario to investigate the potential of CC in human verbal communication: individuals listened to a public announcement and had to reconstruct the announcement alone or in groups. This situation is frequently encountered by humans in their daily life, for instance at train stations or airports.

Communication analysis is challenging but recent developments of sophisticated methods in machine speech recognition have provided us with powerful tools that allow the analysis of syntax and semantics of human language (Jurafsky and Martin, 2009). Here we apply these novel tools to study if human groups can decrease error rate (syntax) and increase semantic understanding compared to single individuals in an everyday task. In this study we particularly focused on the question whether groups can outperform their best member.

Material and methods

Experimental setup

We recruited 167 student volunteers from the University of Bielefeld (Germany) participating in a course on behavioural ecology (April 2011). Participants were divided in 21 groups. All groups consisted of eight members, except one group which had 7 members. Informed consent was obtained from all participants and data collection was anonymous. Prior to the experiment we communicated to all participants that they were allowed to leave at any time. All procedures were carried out in accordance with the Declaration of Helsinki. We deemed it unnecessary to apply for formal ethical approval for this study as it is highly unlikely that participants would feel uncomfortable in participating in this simple and straightforward task. Listening to a sentence and reconstructing a sentence is a very simple task that most people perform on a daily basis without any negative consequences. Moreover, the experiment was part of a student practical for which no ethical approval was required and the students used the data afterwards for learning about experimental design.

In the experiment, two sentences in German and of equal length were played back to the participants (See Table 5.1). These were announcements that are typically audible at a train station or an airport. We added echo, white noise and a 55 Hz tone to both sentences using Audacity (http://audacity.sourceforge.net) mimicking a real-life situation at a public place as for instance encountered at a train station or an airport.

All groups underwent two treatments: “individual treatment” and “group treatment”. In both treatments, participants listened first to a sentence and were given 1 minute to individually write down the sentence as they heard it. Then, for the individual treatment, participants were permitted four additional minutes to improve their sentences individually. In the group treatment, the participants had four minutes to discuss and write down one consensual sentence. All groups received each sentence once (i.e., one sentence during the individual treatment and the other during the group treatment). The order of the two treatments (i.e., individual or group) and the two sentences were randomized so that each of the 4 combinations was performed with approximately the same number of groups (5 or 6).

The participants wrote their sentences on sheets that had 30 boxes and were asked to write one word per box and to leave blanks where they thought that a word was missing. The number of boxes far exceeded the actual number of words in the sentences

to avoid limiting the participants or giving them a clue regarding the actual number of words.

Analysis

The quality of the reconstructed sentences was evaluated on two levels: syntactic, i.e. regarding the correctness of the word sequences, and semantic, i.e. regarding the correctness of the pieces of information contained in the sentences.

For the syntactic analysis we used the "Word Error Rate" (WER), which is the standard evaluation metric for speech recognition (Jurafsky and Martin, 2009). The WER is the minimum number of changes (insertions, deletions, and substitutions of words) needed to transform the correct sentence into the reconstructed one, divided by the number of words in the correct sentence (see Table 5.2 for an example). The WER was calculated using the Speech Recognition Scoring Toolkit (version 2.3.5) of the National Institute of Standards and Technology (http://www.nist.gov/itl/iad/mig/tools.cfm).

The meaning of a sentence with a low WER is not necessarily more correct than that of a sentence with a high WER because the WER does not take the semantic relevance of words into account. Therefore, in our semantic analysis we looked at particular pieces of information, called "items" that constituted the meaning of the sentences. We identified 8 items for the train station announcement and 7 items for the airport announcement (See Table 5.1). We evaluated the semantic correctness using the following measures that are widely used in the fields of information retrieval and speech recognition (Jurafsky and Martin, 2009).

Precision (P) = *Number of correct items in the reconstructed sentence / Total number of items in the reconstructed sentence*

Recall (R) = *Number of correct items in the reconstructed sentence / Total number of items in the correct sentence*

The precision measures the degree to which the retrieved information is correct. The recall measures how much of the available information was successfully retrieved. To evaluate the overall quality of information retrieval, it is common to combine them by computing their harmonic mean, called the F-measure (Jurafsky and Martin, 2009):

$F = 2PR / (P+R)$

The F-measure was calculated using an own script.

Decision mechanism

To understand how groups arrived at communal decisions we studied the transition from the collection of independent responses to the group response during the "group treatment". For this we listed all eight independent responses per word per group and compared this to the group decision of that particular word and that particular group. We distinguished between the categories: consensus (i.e., all independent responses that were given were identical to the group response), majority (i.e., the group response corresponded to the word that was most often reconstructed during the independent responses), tie (i.e., the group response corresponded to one of two (or more) words that were most often reconstructed during the independent responses), minority (i.e., the group response was present in the independent responses but was not one of the words that were most often given in the independent responses) and invented (i.e., the group response was not present in the independent responses). We studied how frequently these different situations occurred and whether they led to better decisions. For this we calculated the rate of correct responses for the independent responses (varying between 0 and 1) and compared this to the group decision (either 0 or 1) (hereafter called: "success rate"). Whenever individuals had no answer for a particular word (i.e., did not hear it) we treated this as "incorrect".

We also studied the group performance as compared to the combination of the best responses from all independent responses (i.e., combining the best answers of all given independent opinions). This allowed us to study if there was a so-called assembly bonus effect present which means that group performance is better than the performance of all individual group members or any combination of individual member efforts (Collins and Guetzkow, 1964; Tindale and Larson, 1992).

Statistics

For the individual treatment, we calculated the WER and F-measure of all individuals and of the best individual after one minute and after four additional minutes. For the group treatment, we calculated the WER and F-measure of the best individual after 1 minute

and the group performance (i.e., after four minutes of discussion). A direct comparison between the best individual after four additional minutes in the individual treatment and the group consensus after four minutes in the group treatment was not possible since we found strong effects of both treatment order and sentence (see below) preventing a direct comparison. Therefore, we ran a separate analysis within each treatment (i.e., individual and group) to quantify the effect of additional minutes on individual and group performance. The success of reconstructing the two sentences (measured by WER and F-measure) by (1) all members, (2) best members and (3) groups were analysed using (separate) generalized linear mixed models with a logit link function (glmmPQL function in package MASS in R, version 2.14.1). As fixed effects in all models we included time (i.e., 1 or 4 minutes) sentence and treatment order (i.e., first or second experiment). For the model including all members, we included individual nested in group as a random effect. For the models including best members or groups, we included group as random effect.

Results

Individual treatment

In the individual treatment, there was no difference in WER or F-measure after 1 or 4 minutes including all individuals (Fig. 5.1a; Table 5.3). There was an effect of treatment order on WER and F-measure during the individual treatment (Table 5.3). Participants that started with the individual treatment performed worse during the individual treatment (i.e., higher WER and lower F-measure) than those that finished with the individual treatment. Additionally, there was an effect of sentence on F-measure but not on WER (Table 5.3). Likewise, the best individuals of the group did not improve in WER or F-measure with additional time (WER: 1 minute: (mean ± SD=) 0.41 ± 0.09, 4 minutes: 0.37 ± 0.10, P = 0.23; F-measure: 1 minute: 0.68 ± 0.09; 4 minutes: 0.67 ± 0.09, P = 0.45; Fig. 5.1b).

Group treatment

Groups scored significantly better on both the WER and the F-measure than the best performing individual in the group treatment (WER: 1 minute: (mean ± SD=) 0.37 ± 0.09, 4 minutes: 0.29 ± 0.09; F-measure: 1 minute: 0.69 ± 0.09; 4 minutes: 0.75 ± 0.13; Table

5.4, Fig. 5.1c). There was no effect of treatment order or sentence on WER or F-measure during the group treatment (Table 5.4).

Decision mechanism

Most of the group decisions when choosing individual words were based on consensus (n = 178) or majority (n = 129), followed by ties (n = 51) and minority (n = 11). A few words (n = 6) were not present in the individual responses but were invented (Fig. 5.2). The success rate of groups was higher than individual success rate in all categories, except during minority voting (Fig. 5.2, see also Discussion).

The combination of best individual responses was significantly better than the group response (WER: 0.16 ± 0.08, $P < 0.01$; F-measure: 0.86 ± 0.09, $P < 0.01$).

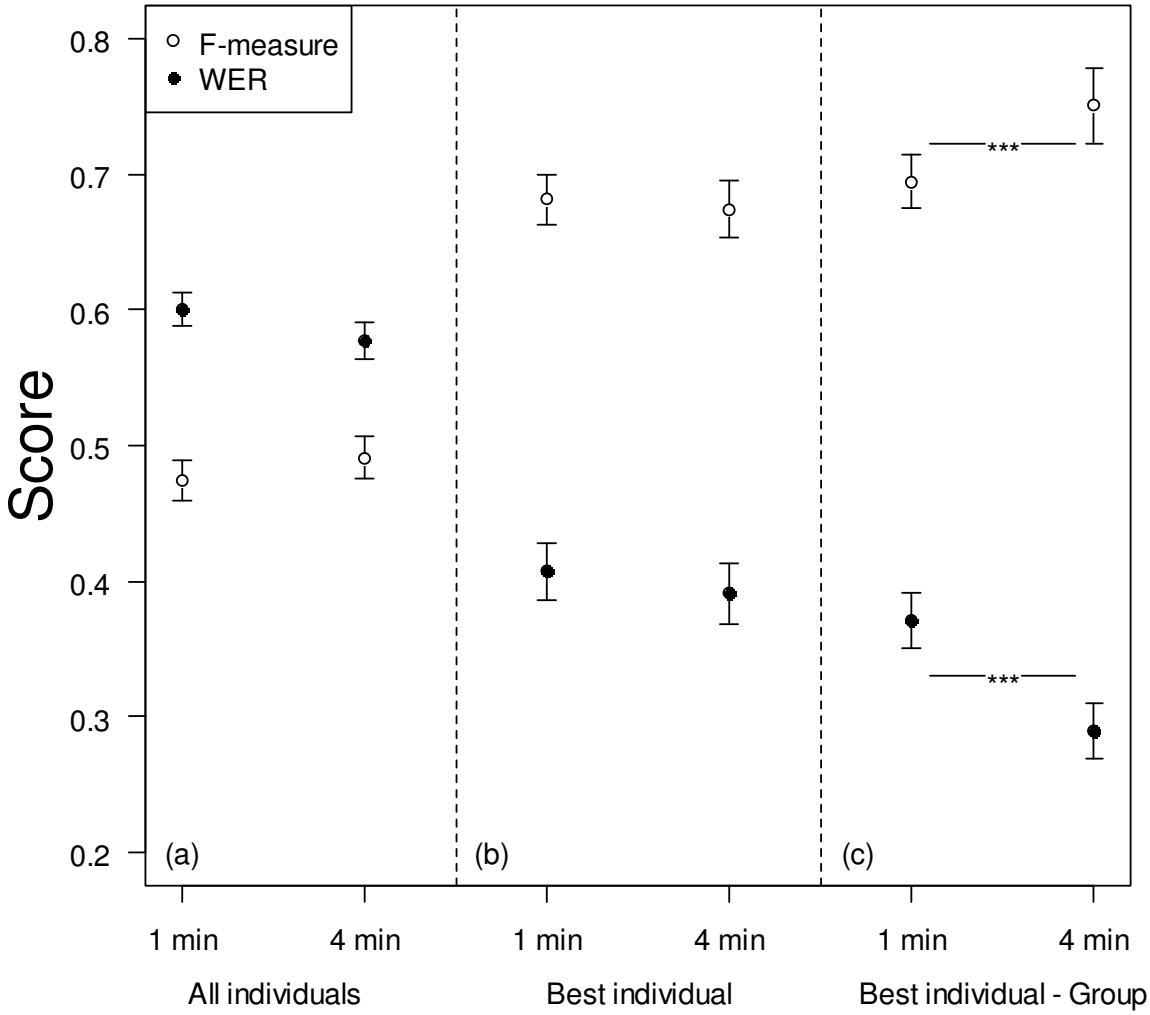

Figure 5.1: Groups outperformed their best members. During the "individual treatment", (a) individuals did not improve their WER or F-measure with extra time. (b) Likewise, the best individuals of each group did not improve with extra time. During the "group treatment", (c) groups had a lower WER and a higher F-measure than the best individuals. Shown are mean ± SE of WER (closed circles) and F-measure (open circles). Data are based on all sentences.

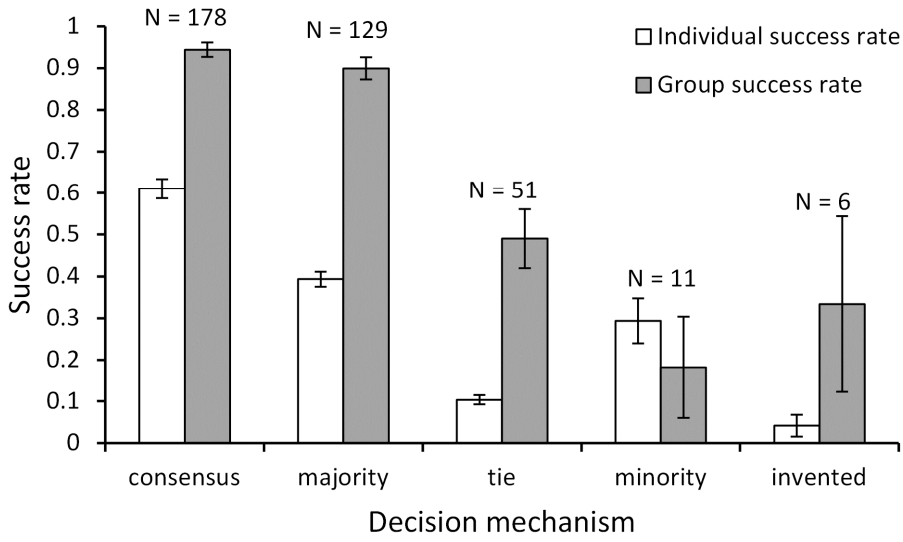

Figure 5.2: Mechanisms of group decisions. Shown are the different categories of how groups decided on a word based on the collection of independent responses/words. Consensus: all independent responses were identical to the group response; majority: the group response corresponded to the word that was most often reconstructed during the independent responses; tie: the group response corresponded to one of two (or more) words that were most often reconstructed during the independent responses; minority: the group response was present in the independent responses but was not one of the words that were most often given in the independent responses; invented: the group response was not present in the independent responses. Per category, the frequency (labelled as "N") and the success rate (mean ± SE) of individuals (white bars) and groups (dark bars) are shown. Whenever an individual did not fill in a word as an independent response this was considered as "incorrect". Majority decisions resulted in higher success rate and were much more frequent than minority decisions, which did not improve success rate.

Table 5.1. Overview of the sentences as used in this study. The original German sentences, the English translation of the sentences and the semantic "items" of both sentences.

Train station announcement	**Airport announcement**
Original sentence: Der Zug aus Reinfeld mit Weiterfahrt nach Hamburg-Dammtor, Abfahrt um 15 Uhr 32, fährt heute auf Gleis 19 ein.	Original sentence: Die Fluggäste des Fluges LG 327 nach Stettin werden gebeten, sich umgehend zum Flugsteig C 31 zu begeben.
Translated sentence: The train from Reinfeld continuing to Hamburg-Dammtor, leaving at 15:32 arrives today on platform 19.	Translated sentence: The passengers of flight LG 327 to Stettin are requested to go to gate C 31 immediately.
List of 8 semantic items: - the subject (train) - the origin of the train (from Reinfeld) - the destination of the train (to Hamburg Dammtor) - the fact that the announcement is about a departure - the time of departure (15:32) - the action of the train (arrives) - the date (today) - the platform (platform 19)	List of 7 semantic items: - the addressees (passengers) - the airline code (LG) - the flight number (327) - the destination of the flight (to Stettin) - where to go (gate) - the gate number (C 31) - the requested action (go to)

Table 5.2. Example for the computation of the Word Error Rate (WER). The first sentence is the correct one, the second sentence is the reconstructed one. The insertions, deletions, and substitutions are marked by 'I', 'D', and 'S', respectively. For this example the WER = (number of changes) / (number of words in the correct sentence) = 4/13.

The	train	to	London		is	delayed	for	fifteen	minutes	due	to	bad	weather
The	train	to	London	Euston	is	delayed	for	fifty	minutes	due	to		
				I				S				D	D

Table 5.3. Result of the generalized linear mixed model analysis of the "individual treatment" including all individuals. Shown are the effects of time (i.e., performance after 1 minute or after 4 additional minutes), order (i.e., sentence being played first or second) and sentence (i.e., train station or airport announcement) on (a) the Word Error Rate and (b) the F-measure.

(a) Word error rate (WER)				
	estimate	Std. Error	t	*P*
(Intercept)	0.858	0.113	7.597	< 0.001
Time	-0.100	0.066	-1.514	0.131
Order	-0.627	0.121	-5.167	< 0.001
Sentence	-0.254	0.121	-2.091	0.051
(b) F-measure				
	estimate	Std. Error	t	*P*
(Intercept)	-0.737	0.120	-6.125	< 0.001
Time	0.070	0.076	0.918	0.359
Order	0.622	0.128	4.850	< 0.001
Sentence	0.626	0.128	4.874	< 0.001

Table 5.4. Results of the generalized linear mixed model analysis of the "group treatment". Shown are the effects of time (i.e., performance of the best individual after 1 minute or the group decision after 4 additional minutes), order (i.e., sentence being played first or last) and sentence (i.e., train station or airport announcement) on (a) the Word Error Rate and (b) the F-measure.

(a) Word error rate (WER)				
	estimate	Std. Error	t	*P*
(Intercept)	-0.754	0.145	-5.218	< 0.001
Time	-0.372	0.101	-3.670	0.002
Order	0.174	0.153	1.133	0.272
Sentence	0.254	0.154	1.655	0.115
(b) F-measure				
	estimate	Std. Error	t	*P*
(Intercept)	1.087	0.201	5.399	< 0.001
Time	0.288	0.126	2.282	0.034
Order	-0.133	0.217	-0.614	0.547
Sentence	-0.341	0.217	-1.571	0.134

Discussion

We show that groups were able to decrease the number of errors and increase the semantic value of reconstructed sentences in a realistic context. In the individual treatment, extra time did not improve the performance, whereas in the group treatment the group outcome was better than the single best individual (Katsikopoulos and King, 2010; Kerr and Tindale, 2004).

Groups performed better than their best individual both at the syntactic level (WER) and at the semantic level (F-measure). The developments in machine speech recognition and collective cognition have so far been separate fields of research. We believe that tools developed in machine speech recognition can open up new possibilities to study how language is used and processed by human groups. This can increase our understanding of how and why human groups use language. This is an important consideration, since language is thought to have played a critical role in the evolution of hominids (Noble and Davidson, 1996) giving them a unique way of sharing information. Due to our limited number of groups (21) and sentences (2) further studies are, however, necessary to evaluate the robustness of our findings. An important consideration is how group improvement is affected by the complexity of the sentence in terms of syntax and semantics. Does group improvement occur only at a narrow range of complexity, or at a broad spectrum of complexity levels? Also further research is warranted to understand how group improvement in sentence reconstruction tasks scales with group size (see also Sorkin, Hays, & West, 2001).

In many previous studies on CC in humans, information is aggregated computationally by the experimenter, post-hoc (King et al., 2011a; Krause et al., 2011; Lorenz et al., 2011; but see King et al., 2011b). Here we obtained independent information from the study subjects, who were then allowed to communicate. The aggregation was thus done by the subjects, simulating real group decision-making in humans. This communication is a key ingredient for CC to arise. It allows participants to exchange not only their opinions but also their level of confidence, a critical piece of information (Bahrami et al., 2010) enabling others to judge how relevant the separate pieces of information are. This allows groups to make better decisions than individuals even in the absence of feedback on individual performances (Bahrami et al., 2010). It would be interesting to see how well groups would do in the absence of communication and only show individuals the opinions of their group members. This would allow

quantification of the importance of the communication aspect. For simple tasks such as estimating quantities, and provided that estimations are independent and then aggregated, group size is one of the main predictors of decision accuracy (Krause et al., 2011b). However, for more complex tasks (such as sentence reconstructions) a benefit of CC with increasing group size is not always a given since larger groups might face communication difficulties. Moreover, CC is not suitable for all types of problems and in some cases it is better to follow the expert (Katsikopoulos and King, 2010; Krause et al., 2011b). Group discussions can even impair decision accuracy due to the inequality of individual influence (Sunstein, 2005) because the opinion of others can negatively influence individual decisions (Lorenz et al., 2011).

Majority decisions occurred much more often than minority decisions (Fig. 5.2) suggesting that there were no strong leaders or dominant individuals present that managed to override majorities (King et al., 2009). During the group discussions, there was often a substantial proportion of individuals (on average 42.3%) that did not fill in a word which might have facilitated majority decisions since the presence of uninformed individuals can increase democratic, majority decisions (Couzin et al., 2011). Majority decisions led to much better decisions, as opposed to minority decisions which deteriorated decisions, illustrating that majority decision is a successful strategy in sentence reconstruction tasks. Most other types of decisions also led to better decisions (Fig. 5.2) and in a few cases groups managed to find the correct word whereas it was not present in their individual responses. This, however, occurred only rarely and we did not find evidence in favour of the assembly bonus effect, which means that group performance is better than the performance of any individual group member or any combination of individual member efforts (Collins and Guetzkow, 1964; Tindale and Larson, 1992). In contrast, the best possible combination of individual responses was significantly better than the group response indicating that although correct words were available, the groups were not always able to incorporate them into their final answer.

Ultimately, the critical test of collective cognition is the actual decision that groups and individuals would make (i.e., would they have caught the train or plane?). We did not study decision accuracy directly but extracted meaning from reconstructed sentences. Evaluating the pragmatic level is a difficult theoretical problem to solve (Jurafsky and Martin, 2009). One possibility is to ask people to carry out the task. However, if it is an everyday problem, people might not only use collective cognition but start using other strategies as well, such as looking at a map or asking professionals. And

if on the other hand the problem is too limited and artificial, then the result would not be representative.

From an evolutionary point of view, the fact that groups beat even their best members shows that not only the average individuals, but also the top ones have an incentive to join a group to solve complex problems. Assessing the costs and benefits (to arrive at fitness measures) of such strategies (i.e. solving a problem alone or as part of a group) remains an important challenge for future studies.

6. General discussion

Collective cognition is widespread in the animal kingdom and certain properties of collective cognition have even been described in plants (Baluska et al., 2010) and cancer cells (Deisboeck and Couzin, 2009). This thesis investigated several aspects of collective cognition encountered in fish and in humans. I will now sum up the results described in the previous chapters and discuss them in a broader context.

In this thesis, I showed that for problems as diverse as estimating a quantity (chapter 2), distinguishing between edible and non-edible items (chapter 3), deciding to escape or stay depending on a perceived threat (chapter 4), and deciphering a noisy message (chapter 5), decisions improve with group size, confirming previous studies on various species (Laughlin et al., 2003; Liker and Bókony, 2009; Ward et al., 2008, 2011). I showed this improvement for groups of humans (chapters 2, 4, 5) by comparing groups to individuals, and for guppies in various group sizes (chapter 3). This improvement can result from different types of interactions: pure aggregation (chapter 2), visual observation (chapters 3 and 4), or discussion (chapter 5). Some of these interactions such as visual observation and reaction to other individuals' movement were observed in guppies (chapter 3) as well as in humans (chapter 4), and have been reported in various species (Conradt and Roper, 2003; Cresswell, 1994; Lima, 1994). Certain mechanisms leading to collective decisions such as quorum responses have been observed in humans (chapter 4) and in animals such as social insects (Seeley and Visscher, 2004; Sumpter and Pratt, 2009), primates (Sueur et al., 2010), and fish (Ward et al., 2008). This suggests that the processes underlying collective decisions via movement are remarkably similar between humans and animals.

Furthermore, in the chapters 2 and 5, I showed that, in addition to improving decision compared to average individuals, groups are able to beat their best members in different contexts. I showed that groups can outperform their best member if a given problem occurs repeatedly (chapter 2), or if the task is very complex (chapter 5). In the wild, these cognitive problems are analogous to detecting a frequent predator that is slightly cryptic, or a rarer but extremely cryptic predator. In both cases, the stakes can be high enough for even the best members to benefit from joining a group.

Studying collective cognition in the wild versus in simulation experiments

Throughout this thesis, I have looked at collective cognition from both ends of the spectrum: from wild conditions in free-ranging guppies (chapter 3) to very controlled simulation experiments in humans (chapters 2, 4, and 5).

When studying collective behaviour, groups are the unit of replication. The high number of individuals needed therefore adds to the inherent difficulties of field experiments. This explains why the number of studies that have investigated collective cognition in the wild is limited (Ballerini et al., 2008; Dell'Ariccia et al., 2008; Kenward, 1978; Treherne and Foster, 1981) and most studies on collective decision-making have been done in captivity (Amé et al., 2006; Franks et al., 2006; Herbert-Read et al., 2011; Ward et al., 2011) or on humans (Dyer et al., 2008; Faria et al., 2009; King et al., 2011b; Kurvers et al., 2014a; Wolf et al., 2013). Although this allows studying the function and mechanisms of collective cognition in very controlled conditions, those conditions may also be oversimplified or even unnatural and omit important factors. Animals live in a complex environment and are subject to many influences such as inter- and intraspecific competition, predation, weather fluctuations, etc. Some of these factors may directly influence collective cognition. Studies in natural conditions are thus necessary to verify that the effects predicted by models and observed in laboratory studies are still robust in the wild.

In chapter 3, I studied the ability of wild guppies to make better decisions as a function of group size. This investigation followed a comparative approach and took place at locations that are subject to different levels of predation. Previous studies have shown that predation can affect social structure and behaviour (Kelley et al., 2011) but few studies have investigated collective cognition in animals across populations that are subject to different predation pressures in the wild (but see Webster & Laland, 2008 for a laboratory experiment). I showed that guppies indeed made better decisions with increasing group size, most likely as a result of social information use. Although guppies were less active in high predation sites, guppies' decision quality did not differ between high and low predation sites. It would be interesting for future studies to look into the dynamics of how information is transmitted.

In the chapters 2, 4, and 5, I investigated collective cognition in human groups, exposing them to tasks that increased in complexity and testing different rules for

combining information, ranging from aggregation of independent decisions, to group discussions. Working with humans facilitated the separation between private and social information.

Through the use of questionnaires (chapters 2 and 5) and voting machines (chapter 4), it was possible to identify each individual's private information before it got shared with other group members. Furthermore, it was possible to manipulate the level of interaction between the participants: i) no interaction (chapter 2) where private estimates were aggregated; ii) movement behaviour without verbal communication, mimicking situations occurring in wild animals (chapter 4); iii) open discussion (chapter 5), which is a typical process for groups of humans to reach a decision.

In chapter 2, participants were asked to estimate a quantity several times. The average of their guesses was then used as a proxy to estimate the potential of collective cognition. I found that in this context of a repeated task, some individuals were able to beat the group several times, but not every time. In this experiment, the best individual could beat the group 7 times out of 10. This showed that even for simple tasks, top performers can benefit from being in a group if these tasks are encountered repeatedly. Using an aggregation rule may not seem very realistic when comparing to animal groups in the wild, but this allowed me to look into the potential for collective cognition without having to worry about the potential weaknesses of interactions.

In chapter 4, participants were asked to escape or stay depending on the presence or absence of a simulated predator. They were only allowed to move forward or backward but not to verbally communicate. Despite its apparent simplicity, this setup has a lot in common with many wild animal groups since in many situations the main form of communication for this type of decision is through making escape movements that can indicate the presence of a predator (Beauchamp and Ruxton, 2007; Conradt and Roper, 2003; Sueur et al., 2010; Ward et al., 2011). Even in humans, observing the movement behaviour of others can elicit a movement response (Faria et al., 2010a). I found that individuals in a group were able to improve their decision using social information. Compared to individuals alone, individuals in groups were able to increase their true positive rates and decrease their false positive rates simultaneously, using only the information provided by the movement of others. This improvement arose in the absence of any verbal communication. Seeing a given number of individuals in the group escaping was sufficient to trigger an escape by most of the other group members. Individuals used specific threshold rules, also known as quorum rules, which allowed them to make the

correct decision most of the time, following the decision of a fraction of the group that was much smaller than the majority, indicating that the group can benefit from following a minority (Conradt and Roper, 2003; Couzin et al., 2005; Dyer et al., 2008).

In chapter 5, I included verbal discussion as a means for the group to improve decision accuracy. In this experiment, participants had to listen to a distorted audio announcement and, after writing down what they understood from the announcement, they had to discuss with their group members on how to reconstruct the original announcement. I found that groups' results were not only better than those of average individuals but also better than their best members. This showed that when a problem is complex enough, best members can benefit from joining a group. However, high performing individuals still had an influence on the group performance. This problem simulated a situation that humans encounter regularly in everyday life.

Collective cognition to solve increasingly complex tasks

In chapter 2, I studied whether groups are able to outperform not only average individuals, but also the best individuals in the group. Using a simple estimation task, I showed that when the task is repeated often enough, even the best individuals don't manage to beat the group. This confirms theoretical predictions that experts are likely to beat groups in single-shot decisions but that groups perform better for repeated tasks (Katsikopoulos and King, 2010). The task was very simple and the correct solution was unambiguous (the exact number of dots was known). For this task, using a simple aggregation rule was sufficient in order to reap the benefits of collective cognition.

In chapter 3 and 4, I showed how individuals increased their decision accuracy when facing binary choices, using social information based on others' reaction to the stimuli. It has been shown that even without active signalling or individual recognition, a fraction of knowledgeable individuals is able to lead a group (Couzin et al., 2005; Dyer et al., 2008), provided that group members have the incentive to stay together. A similar mechanism has been shown in golden shiners (*Notemigonus chrysoleucas*) in an experiment in which parts of a tank were randomly shaded or illuminated (Berdahl et al., 2013). Golden shiners prefer darker areas. The study showed that individuals that were in shaded areas swam slower than those in illuminated areas, resulting in the whole group turning towards the shaded area.

It has been shown however, that some problems require expert knowledge (Krause et al., 2010) and in such cases simply aggregating individual decisions can deteriorate collective decisions. In such cases, the group might not beat the expert, but all its members might benefit from the knowledge shared by the expert. This is known as the "pool of competence hypothesis" (Wolf and Krause, 2014). In such cases, reaching a consensus decision by discussion can be a better strategy than a simple polling of every opinion. Discussion allows indeed to get rid of outliers and a knowledgeable expert can expose and explain why a given solution is best. I showed in chapter 5 that when the problem to solve is more complex, individuals also benefit from being part of a group. The ability of groups to beat even the best individuals has far-reaching implications in human health, as doctors are better as a group at diagnosing cancer than even the best of them (Kurvers et al., 2015; Wolf et al., 2015).

The complexity of the experiment presented in chapter 5 reflects better the complexity of decisions that humans need to make on a regular basis and discussion may be the best way for the moment to make decisions on complex issues. However, discussion can be influenced by many factors such as leadership (or at least hierarchy), differences in information, personality (for example differences in aggressiveness or along the boldness/shyness continuum), and care needs to be taken on how to best structure a discussion. In many cases, some information can be lost and the final outcome may be worse than if obtained by other aggregation methods, and also worse than at least a portion of the group members (Janis, 1982; Lorenz et al., 2011).

Effect of personality and network structure

Chapter 2 used a simple aggregation, leading to a very democratic decision, preventing any effect of personality or network position, or potential bias due to age, sex, race, etc. However the studies presented in chapters 3, 4, and 5 could be influenced by such factors. The study presented in chapter 4 shows that movement behaviour without vocal communication is enough for groups to benefit from their members. It would be interesting to have more information about personality differences that affect the group. For example, leadership can be associated with boldness (Leblond and Reebs, 2006). However, a bold individual would probably not be followed by the rest of the group if most of them know about the presence of a predator (Ward et al., 2011). So boldness may not always be a predictor of who the leader is. In situations of exploring for new

territories or food, boldness/exploration may be associated with leadership. But in case of a costly threat, bold leaders are probably not blindly followed. Furthermore, social responsiveness, and therefore the use of social information may vary between different personalities (Wolf and Krause, 2014).

It is also important to understand the underlying structure of the group. This can be done using social network analysis. As not every group uses completely democratic or completely despotic decision-making, the network structure is likely to influence the process leading to a group decision. This necessity is illustrated by the "majority illusion" (Lerman et al., 2015), in which each individual from a group is under the impression that a majority of his companions is in favour of one option, whereas in truth, this option is favoured by a minority. This phenomenon is due to the minority members being very well connected compared to the majority members.

Outlook and avenues for future research

I will conclude this chapter with remarks on the future of studies on collective behaviour, especially on collective cognition and decision-making.

Many models try to understand how individuals make decisions, from very simple models based on heuristics (Gigerenzer and Gaissmaier, 2011) to complex models delineating the use of social information versus private information (Arganda et al., 2012). A deeper understanding of the social structure of a group, using social network analysis would help understanding the processes underlying collective cognition (Kurvers et al., 2014b). Furthermore, we have seen that personality as well can influence group structure and should be taken into account when studying group decision-making (Wolf and Krause, 2014).

Recent advances in neuroimaging enabling to examine in real time the firing of neurons in the brain of live organisms are promising tools to investigate the underlying processes at the brain level during decision-making (Ahrens et al., 2013). Fast developments in the field of robotics will allow to test collective decision-making models either by integrating robots within groups of animals, or by using autonomous robots capable of interacting with each other (Faria et al., 2010b; Knight, 2005; Krause et al., 2011a). Finally, technologies such as global positioning systems (GPS), acoustic tags or radio-frequency identification (RFID) are becoming increasingly smaller and affordable, bringing the possibility to track the positions and interactions between numerous

individuals and allowing reality mining of animal groups in the wild (Krause et al., 2013). These techniques are already being used to study decision-making in primates (King and Sueur, 2011), pigeons (Dell'Ariccia et al., 2008), ants (Robinson et al., 2009), and sharks (Guttridge et al., 2010).

Beyond understanding how animal groups make collective decisions, understanding how humans can improve their decision-making by using swarm intelligence is a crucial issue with many applications susceptible to affect our lives, in various domains ranging from medicine (Kurvers et al., 2015; Wolf et al., 2015) to natural resources management (Arlinghaus and Krause, 2013).

References

Ahrens, M.B., Orger, M.B., Robson, D.N., Li, J.M., and Keller, P.J. (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. Nat. Methods *10*, 413–420.

Amé, J.M., Halloy, J., Rivault, C., Detrain, C., and Deneubourg, J.L. (2006). Collegial decision making based on social amplification leads to optimal group formation. Proc. Natl. Acad. Sci. U. S. A. *103*, 5835–5840.

Arganda, S., Pérez-Escudero, A., and Polavieja, G.G. de (2012). A common rule for decision making in animal collectives across species. Proc. Natl. Acad. Sci. U. S. A. *109*, 20508–20513.

Arlinghaus, R., and Krause, J. (2013). Wisdom of the crowd and natural resource management. Trends Ecol. Evol. *28*, 8–11.

Armstrong, J.S. (2001). Combining forecasts. Princ. Forecast. Handb. Res. Pract. 417–439.

Arrow, K.J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J.O., Levmore, S., Litan, R., Milgrom, P., Nelson, F.D., et al. (2008). The promise of prediction markets. Science *320*, 877–878.

Bahrami, B., Olsen, K., Latham, P.E., Roepstorff, A., Rees, G., and Frith, C.D. (2010). Optimally interacting minds. Science *329*, 1081–1085.

Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., and Procaccini, A. (2008). Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. Proc. Natl. Acad. Sci. U. S. A. *105*, 1232–1237.

Baluska, F., Lev-Yadun, S., and Mancuso, S. (2010). Swarm intelligence in plant roots. Trends Ecol. Evol. *25*, 682–683.

Beauchamp, G. (2010). Determinants of false alarms in staging flocks of semipalmated sandpipers. Behav. Ecol. *21*, 584–587.

Beauchamp, G., and Ruxton, G.D. (2007). False alarms and the evolution of antipredator vigilance. Anim. Behav. *74*, 1199–1206.

Bednekoff, P.A., and Lima, S.L. (1998). Re-examining safety in numbers: interactions between risk dilution and collective detection depend upon predator targeting behaviour. Proc. R. Soc. B Biol. Sci. *265*, 2021–2026.

Bell, A.M., Hankison, S.J., and Laskowski, K.L. (2009). The repeatability of behaviour: a meta-analysis. Anim. Behav. *77*, 771–783.

Beni, G., and Wang, J. (1989). Swarm Intelligence in Cellular Robotic Systems. In Proceedings of NATO Advanced Workshop on Robots and Biological Systems, (Tuscany, Italy), pp. 703–712.

Berdahl, A., Torney, C.J., Ioannou, C.C., Faria, J.J., and Couzin, I.D. (2013). Emergent sensing of complex environments by mobile animal groups. Science *339*, 574–576.

van Bergen, Y., Coolen, I., and Laland, K.N. (2004). Nine-spined sticklebacks exploit the most reliable source when public and private information conflict. Proc. R. Soc. B Biol. Sci. *271*, 957–962.

Bergman, G., and Donner, K.O. (1964). An analysis of the spring migration of the common scoter and the long-tailed duck in Southern Finland. Acta Zool. Fenn. *105*, 1–59.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. J. Polit. Econ. *100*, 992–1026.

Biro, D., Sumpter, D.J.T., Meade, J., and Guilford, T. (2006). From compromise to leadership in pigeon homing. Curr. Biol. *16*, 2123–2128.

Blumstein, D.T., Evans, C.S., and Daniels, J.C. (2006). JWatcher.

Boland, P.J. (1989). Majority systems and the Condorcet jury theorem. The Statistician *38*, 181–189.

Bonabeau, E., Dorigo, M., and Theraulaz, G. (1999). Swarm Intelligence: From Natural to Artificial Systems (USA: Oxford University Press).

Breder, C.M. (1954). Equations descriptive of fish schools and other animal aggregations. Ecology *35*, 361–370.

Brown, C., and Laland, K.N. (2003). Social learning in fishes: a review. Fish Fish. *4*, 280–288.

Brown, C., Laland, K.N., and Krause, J. (2006). Fish Cognition and Behavior (Oxford, UK: Wiley-Blackwell).

Clément, R.J.G., Krause, S., von Engelhardt, N., Faria, J.J., Krause, J., and Kurvers, R.H.J.M. (2013). Collective cognition in humans: groups outperform their best members in a sentence reconstruction task. PLoS One *8*, e77943.

Codling, E.A., Pitchford, J.W., and Simpson, S.D. (2007). Group navigation and the "many-wrongs principle" in models of animal movement. Ecology *88*, 1864–1870.

Collins, B.E., and Guetzkow, H.S. (1964). A Social Psychology of Group Processes for Decision-Making (New York: Wiley).

Condorcet, M. de (1785). Essay on the Application of Analysis to the Probability of Majority Decisions. Paris Impr. R.

Conradt, L. (2011). Collective behaviour: When it pays to share decisions. Nature *471*, 40–41.

Conradt, L. (2012). Models in animal collective decision-making: information uncertainty and conflicting preferences. Interface Focus *2*, 226–240.

Conradt, L., and List, C. (2009). Group decisions in humans and animals: a survey. Philos. Trans. R. Soc. B Biol. Sci. *364*, 719–742.

Conradt, L., and Roper, T.J. (2003). Group decision-making in animals. Nature *421*, 155–158.

Conradt, L., and Roper, T.J. (2005). Consensus decision making in animals. Trends Ecol. Evol. *20*, 449–456.

Conradt, L., and Roper, T.J. (2007). Democracy in animals: the evolution of shared group decisions. Proc. R. Soc. B Biol. Sci. *274*, 2317–2326.

Conradt, L., and Roper, T.J. (2009). Conflicts of interest and the evolution of decision sharing. Philos. Trans. R. Soc. B Biol. Sci. *364*, 807–819.

Conradt, L., and Roper, T.J. (2010). Deciding group movements: Where and when to go. Behav. Processes *84*, 675–677.

Conradt, L., Krause, J., Couzin, I.D., and Roper, T.J. (2009). "Leading according to need" in self-organizing groups. Am. Nat. *173*, 304–312.

Coolen, I., Bergen, Y.V., Day, R.L., and Laland, K.N. (2003). Species difference in adaptive use of public information in sticklebacks. Proc. R. Soc. Lond. B Biol. Sci. *270*, 2413–2419.

Coolen, I., Ward, A.J.W., Hart, P.J.B., and Laland, K.N. (2005). Foraging nine-spined sticklebacks prefer to rely on public information over simpler social cues. Behav. Ecol. *16*, 865.

Couzin, I.D. (2009). Collective cognition in animal groups. Trends Cogn. Sci. *13*, 36–43.

Couzin, I.D., Krause, J., Franks, N.R., and Levin, S.A. (2005). Effective leadership and decision-making in animal groups on the move. Nature *433*, 513–516.

Couzin, I.D., Ioannou, C.C., Demirel, G., Gross, T., Torney, C.J., Hartnett, A., Conradt, L., Levin, S.A., and Leonard, N.E. (2011). Uninformed individuals promote democratic consensus in animal groups. Science *334*, 1578–1580.

Cresswell, W. (1994). Flocking is an effective anti-predation strategy in redshanks, *Tringa totanus*. Anim. Behav. *47*, 433–442.

Cresswell, W., Hilton, G.M., and Ruxton, G.D. (2000). Evidence for a rule governing the avoidance of superfluous escape flights. Proc. R. Soc. B Biol. Sci. *267*, 733–737.

Croft, D.P., Morrell, L.J., Wade, A.S., Piyapong, C., Ioannou, C.C., Dyer, J.R.G., Chapman, B.B., Wong, Y., and Krause, J. (2006). Predation risk as a driving force for sexual segregation: a cross-population comparison. Am. Nat. *167*, 867–878.

Dall, S.R.X., Houston, A.I., and McNamara, J.M. (2004). The behavioural ecology of personality: consistent individual differences from an adaptive perspective. Ecol. Lett. *7*, 734–739.

Danchin, E., Giraldeau, L.A., Valone, T.J., and Wagner, R.H. (2004). Public information: From nosy neighbors to cultural evolution. Science *305*, 487–491.

Davies, N.B., Krebs, J.R., and West, S.A. (2012). An introduction to behavioural ecology (UK: Wiley-Blackwell).

Deary, I., J., Penke, L., and Johnson, W. (2010). The neuroscience of human intelligence differences. Nat. Rev. Neurosci. *11*, 201–211.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: a logarithmic mental number line. Trends Cogn. Sci. *7*, 145–147.

Deisboeck, T.S., and Couzin, I.D. (2009). Collective behavior in cancer cell populations. BioEssays *31*, 190–197.

Dell'Ariccia, G., Dell'Omo, G., Wolfer, D.P., and Lipp, H.P. (2008). Flock flying improves pigeons' homing: GPS track analysis of individual flyers versus small groups. Anim. Behav. *76*, 1165–1172.

Devereux, C.L., Whittingham, M.J., Fernandez-Juricic, E., Vickery, J.A., and Krebs, J.R. (2006). Predator detection and avoidance by starlings under differing scenarios of predation risk. Behav. Ecol. *17*, 303–309.

Dorofeev, S., and Grant, P. (2006). Statistics for real-life sample surveys: Non-simple-random samples and weighted data (UK: Cambridge University Press).

Dyer, J.R.G., Ioannou, C.C., Morrell, L.J., Croft, D.P., Couzin, I.D., Waters, D.A., and Krause, J. (2008). Consensus decision making in human crowds. Anim. Behav. *75*, 461–470.

Dyer, J.R.G., Johansson, A., Helbing, D., Couzin, I.D., and Krause, J. (2009). Leadership, consensus decision making and collective behaviour in humans. Philos. Trans. R. Soc. B Biol. Sci. *364*, 781–789.

Eagle, N., and Pentland, A. (2006). Reality mining: sensing complex social systems. Pers. Ubiquitous Comput. *10*, 255–268.

Elgar, M.A. (1989). Predator vigilance and group size in mammals and birds: a critical review of the empirical evidence. Biol. Rev. *64*, 13–33.

Esser, J.K. (1998). Alive and well after 25 years: A review of groupthink research. Organ. Behav. Hum. Decis. Process. *73*, 116–141.

Faria, J.J., Codling, E.A., Dyer, J.R.G., Trillmich, F., and Krause, J. (2009). Navigation in human crowds; testing the many-wrongs principle. Anim. Behav. *78*, 587–591.

Faria, J.J., Krause, S., and Krause, J. (2010a). Collective behavior in road crossing pedestrians: the role of social information. Behav. Ecol. *21*, 1236–1242.

Faria, J.J., Dyer, J.R.G., Clément, R., Couzin, I.D., Holt, N., Ward, A.J.W., Waters, D., and Krause, J. (2010b). A novel method for investigating the collective behaviour of fish: introducing "Robofish." Behav. Ecol. Sociobiol. *64*, 1211–1218.

Fernandez-Juricic, E., Kerr, B., Bednekoff, P.A., and Stephens, D.W. (2004). When are two heads better than one? Visual perception and information transfer affect vigilance coordination in foraging groups. Behav. Ecol. *15*, 898–906.

Fischer, I., and Harvey, N. (1999). Combining forecasts: What information do judges need to outperform the simple average? Int. J. Forecast. *15*, 227–246.

Fox, D.L. (1976). Animal Biochromes and Structural Colours: Physical, Chemical, Distributional & Physiological Features of Coloured Bodies in the Animal World (USA: University of California Press).

Franks, N.R. (1989). Army ants: a collective intelligence. Am. Sci. *77*, 138–145.

Franks, N.R., Dornhaus, A., Best, C.S., and Jones, E.L. (2006). Decision making by small and large house-hunting ant colonies: one size fits all. Anim. Behav. *72*, 611–616.

Galef, B.G. (1991). Information centres of Norway rats: sites for information exchange and information parasitism. Anim. Behav. *41*, 295–301.

Gallup, A.C., Hale, J.J., Sumpter, D.J.T., Garnier, S., Kacelnik, A., Krebs, J.R., and Couzin, I.D. (2012). Visual attention and the acquisition of information in human crowds. Proc. Natl. Acad. Sci. U. S. A. *109*, 7245–7250.

Galton, F. (1907). Vox Populi. Nature *75*, 450–451.

Garnier, S., Gautrais, J., and Theraulaz, G. (2007). The biological principles of swarm intelligence. Swarm Intell. *1*, 3–31.

Gazda, S.K., Connor, R.C., Edgar, R.K., and Cox, F. (2005). A division of labour with role specialization in group–hunting bottlenose dolphins (Tursiops truncatus) off Cedar Key, Florida. Proc. R. Soc. Lond. B Biol. Sci. *272*, 135–140.

Gigerenzer, G., and Gaissmaier, W. (2011). Heuristic decision making. Annu. Rev. Psychol. *62*, 451–482.

Giraldeau, L.A., Valone, T.J., and Templeton, J.J. (2002). Potential disadvantages of using socially acquired information. Philos. Trans. R. Soc. B Biol. Sci. *357*, 1559–1566.

Goss, S., Aron, S., Deneubourg, J.L., and Pasteels, J.M. (1989). Self-organized shortcuts in the Argentine ant. Naturwissenschaften *76*, 579–581.

Grünbaum, D. (1998). Schooling as a strategy for taxis in a noisy environment. Evol. Ecol. *12*, 503–522.

Guttridge, T.L., Gruber, S.H., Krause, J., and Sims, D.W. (2010). Novel acoustic technology for studying free-ranging shark social behaviour by recording individuals' interactions. PLoS ONE *5*, e9324.

Hall, J.R., Cuthill, I.C., Baddeley, R., Shohet, A.J., and Scott-Samuel, N.E. (2013). Camouflage, detection and identification of moving targets. Proc. R. Soc. Lond. B Biol. Sci. *280*, 20130064.

Hamilton, W.J. (1967). Social aspects of bird orientation mechanisms. In Animal Orientation and Navigation, (USA: Oregon State University Press), pp. 57–71.

Hancock, P.A., Milner-Gull, E.J., and Keeling, M.J. (2006). Modelling the many-wrongs principle: The navigational advantages of aggregation in nomadic foragers. J. Theor. Biol. *240*, 302–310.

Handegard, N.O., Boswell, K.M., Ioannou, C.C., Leblanc, S.P., Tjøstheim, D.B., and Couzin, I.D. (2012). The dynamics of coordinated group hunting and collective information transfer among schooling prey. Curr. Biol. *22*, 1213–1217.

Healy, S.D., Bacon, I.E., Haggis, O., Harris, A.P., and Kelley, L.A. (2009). Explanations for variation in cognitive ability: Behavioural ecology meets comparative cognition. Behav. Processes *80*, 288–294.

Herbert-Read, J.E., Perna, A., Mann, R.P., Schaerf, T.M., Sumpter, D.J.T., and Ward, A.J.W. (2011). Inferring the rules of interaction of shoaling fish. Proc. Natl. Acad. Sci. U. S. A. *108*, 18726–18731.

Hilton, G.M., Cresswell, W., and Ruxton, G.D. (1999). Intraflock variation in the speed of escape-flight response on attack by an avian predator. Behav. Ecol. *10*, 391–395.

Hingee, M., and Magrath, R.D. (2009). Flights of fear: a mechanical wing whistle sounds the alarm in a flocking bird. Proc. R. Soc. B Biol. Sci. *276*, 4173–4179.

Hong, L., and Page, S.E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. Proc. Natl. Acad. Sci. U. S. A. *101*, 16385–16389.

Ioannou, C.C., and Krause, J. (2009). Interactions between background matching and motion during visual detection can explain why cryptic animals keep still. Biol. Lett. *5*, 191–193.

Ioannou, C.C., Guttal, V., and Couzin, I.D. (2012). Predatory fish select for coordinated collective motion in virtual prey. Science *337*, 1212–1215.

Janis, I.L. (1971). Groupthink. Psychol. Today Mag. 84–90.

Janis, I.L. (1982). Groupthink: Psychological studies of policy decisions and fiascoes (Boston, MA: Houghton Mifflin).

Jurafsky, D., and Martin, J.H. (2009). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (USA: Prentice Hall).

Katsikopoulos, K.V., and King, A.J. (2010). Swarm intelligence in animal groups: when can a collective out-perform an expert? PLoS One *5*, e15505.

Katz, Y., Tunstrøm, K., Ioannou, C.C., Huepe, C., and Couzin, I.D. (2011). Inferring the structure and dynamics of interactions in schooling fish. Proc. Natl. Acad. Sci. U. S. A. *108*, 18720–18725.

Kelley, J.L., Morrell, L.J., Inskip, C., Krause, J., and Croft, D.P. (2011). Predation risk shapes social networks in fission-fusion populations. PLoS One *6*, e24280.

Kenward, R.E. (1978). Hawks and doves: factors affecting success and selection in goshawk attacks on woodpigeons. J. Anim. Ecol. *47*, 449–460.

Kerr, N.L., and Tindale, R.S. (2004). Group performance and decision making. Annu. Rev. Psychol. *55*, 623–655.

King, A.J., and Cowlishaw, G. (2009). Leaders, followers and group decision-making. Commun. Integr. Biol. *2*, 147–150.

King, A.J., and Sueur, C. (2011). Where next? Group coordination and collective decision making by primates. Int. J. Primatol. *32*, 1245–1267.

King, A.J., Johnson, D.D.P., and Van Vugt, M. (2009). The origins and evolution of leadership. Curr. Biol. *19*, R911–R916.

King, A.J., Cheng, L., Starke, S.D., and Myatt, J.P. (2011a). Is the true "wisdom of the crowd" to copy successful individuals? Biol. Lett. *8*, 197–200.

King, A.J., Narraway, C., Hodgson, L., Weatherill, A., Sommer, V., and Sumner, S. (2011b). Performance of human groups in social foraging: the role of communication in consensus decision making. Biol. Lett. *7*, 237–240.

Knight, J. (2005). When robots go wild. Nature *434*, 954–955.

Kodric-Brown, A. (1989). Dietary carotenoids and male mating success in the guppy: an environmental component to female choice. Behav. Ecol. Sociobiol. *25*, 393–401.

Kolluru, G.R., Grether, G.F., South, S.H., Dunlop, E., Cardinali, A., Liu, L., and Carapiet, A. (2006). The effects of carotenoid and food availability on resistance to a naturally occurring parasite (*Gyrodactylus turnbulli*) in guppies (*Poecilia reticulata*). Biol. J. Linn. Soc. *89*, 301–309.

Koriat, A. (2012). When are two heads better than one and why? Science *336*, 360–362.

Krause, J. (1993). The relationship between foraging and shoal position in a mixed shoal of roach (*Rutilus rutilus*) and chub (*Leuciscus cephalus*): a field study. Oecologia *93*, 356–359.

Krause, J., and Ruxton, G.D. (2002). Living in Groups (Oxford, UK: Oxford University Press).

Krause, J., Ruxton, G.D., and Krause, S. (2010). Swarm intelligence in animals and humans. Trends Ecol. Evol. *25*, 28–34.

Krause, J., Winfield, A.F.T., and Deneubourg, J.-L. (2011a). Interactive robots in experimental biology. Trends Ecol. Evol. *26*, 369–375.

Krause, J., Krause, S., Arlinghaus, R., Psorakis, I., Roberts, S., and Rutz, C. (2013). Reality mining of animal social systems. Trends Ecol. Evol. *28*, 541–551.

Krause, S., James, R., Faria, J.J., Ruxton, G.D., and Krause, J. (2011b). Swarm intelligence in humans: diversity can trump ability. Anim. Behav. *81*, 941–948.

Kurvers, R.H.J.M., Wolf, M., and Krause, J. (2014a). Humans use social information to adjust their quorum thresholds adaptively in a simulated predator detection experiment. Behav. Ecol. Sociobiol. *68*, 449–456.

Kurvers, R.H.J.M., Krause, J., Croft, D.P., Wilson, A.D.M., and Wolf, M. (2014b). The evolutionary and ecological consequences of animal social networks: emerging issues. Trends Ecol. Evol. *29*, 326–335.

Kurvers, R.H.J.M., Krause, J., Argenziano, G., Zalaudek, I., and Wolf, M. (2015). Detection accuracy of collective intelligence assessments for skin cancer diagnosis. JAMA Dermatol. *151*, 1346–1353.

Lassila, K.D. (2008). A brief history of groupthink. Yale Alumni Mag.

Laughlin, P.R., Bonner, B.L., and Miner, A.G. (2002). Groups perform better than the best individuals on letters-to-numbers problems. Organ. Behav. Hum. Decis. Process. *88*, 605–620.

Laughlin, P.R., Zander, M.L., Knievel, E.M., and Tang, T.K. (2003). Groups perform better than the best individuals on letters-to-numbers problems: informative equations and effective strategies. J. Pers. Soc. Psychol. *85*, 684–694.

Laughlin, P.R., Hatch, E.C., Silver, J.S., and Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: effects of group size. J. Pers. Soc. Psychol. *90*, 644–651.

Laughlin, S.B., de Ruyter van Steveninck, R.R., and Anderson, J.C. (1998). The metabolic cost of neural information. Nat. Neurosci. *1*, 36–41.

Lazarus, J. (1979). The early warning function of flocking in birds: An experimental study with captive quelea. Anim. Behav. *27*, 855–865.

Leblond, C., and Reebs, S.G. (2006). Individual leadership and boldness in shoals of golden shiners (Notemigonus crysoleucas). Behaviour *143*, 1263–1280.

Leonard, N.E., Shen, T., Nabet, B., Scardovi, L., Couzin, I.D., and Levin, S.A. (2011). Decision versus compromise for animal groups in motion. Proc. Natl. Acad. Sci. U. S. A. *109*, 227–232.

Lerman, K., Yan, X., and Wu, X.-Z. (2015). The majority illusion in social networks. ArXiv150603022 Phys.

Liker, A., and Bókony, V. (2009). Larger groups are more successful in innovative problem solving in house sparrows. Proc. Natl. Acad. Sci. *106*, 7893–7898.

Lima, S.L. (1994). Collective detection of predatory attack by birds in the absence of alarm signals. J. Avian Biol. *25*, 319–326.

Lima, S.L. (1995a). Back to the basics of anti-predatory vigilance: the group-size effect. Anim. Behav. *49*, 11–20.

Lima, S.L. (1995b). Collective detection of predatory attack by social foragers: fraught with ambiguity? Anim. Behav. *50*, 1097–1108.

List, C. (2004). Democracy in animal groups: a political science perspective. Trends Ecol. Evol. *19*, 168–169.

Lorenz, J., Rauhut, H., Schweitzer, F., and Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. Proc. Natl. Acad. Sci. *108*, 9020–9025.

Macmillan, N.A., and Creelman, C.D. (2004). Detection Theory: A User's Guide (USA: Lawrence Erlbaum Associates).

Magurran, A.E. (2005). Evolutionary Ecology: The Trinidadian Guppy (USA: Oxford University Press).

Mathieu, J., Maynard, M.T., Rapp, T., and Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. J. Manag. *34*, 410–476.

Mattingly, H.T., and Butler, M.J., IV (1994). Laboratory predation on the trinidadian guppy: implications for the size-selective predation hypothesis and guppy life history evolution. Oikos *69*, 54–64.

McNeil, B.J., Keeler, E., and Adelstein, S.J. (1975). Primer on certain elements of medical decision making. N. Engl. J. Med. *293*, 211–215.

Miller, N., Garnier, S., Hartnett, A.T., and Couzin, I.D. (2013). Both information and social cohesion determine collective decisions in animal groups. Proc. Natl. Acad. Sci. U. S. A. *110*, 5263–5268.

Nieder, A., and Dehaene, S. (2009). Representation of Number in the Brain. Annu. Rev. Neurosci. *32*, 185–208.

Noble, W., and Davidson, I. (1996). Human Evolution, Language and Mind: A Psychological and Archaeological Inquiry (UK: Cambridge University Press).

Post, D.J. van der, Weerd, H. de, Verbrugge, Rineke, and Hemelrijk, C.K. (2013). A novel mechanism for a survival advantage of vigilant individuals in groups. Am. Nat. *182*, 682–688.

Pratt, S.C., Mallon, E.B., Sumpter, D.J.T., and Franks, N.R. (2002). Quorum sensing, recruitment, and collective decision-making during colony emigration by the ant Leptothorax albipennis. Behav. Ecol. Sociobiol. *52*, 117–127.

Proctor, C.J., Broom, M., and Ruxton, G.D. (2001). Modelling antipredator vigilance and flight response in group foragers when warning signals are ambiguous. J. Theor. Biol. *211*, 409–417.

Pulliam, H.R. (1973). On the advantages of flocking. J. Theor. Biol. *38*, 419–422.

Radakov, D.V. (1973). Schooling in the Ecology of Fish (Chichester, U.K.: J. Wiley).

R Development Core Team (2015). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing).

Reebs, S.G. (2000). Can a minority of informed leaders determine the foraging movements of a fish shoal? Anim. Behav. *59*, 403–409.

Roberts, G. (1997). How many birds does it take to put a flock to flight? Anim. Behav. *54*, 1517–1522.

Robinson, E.J.H., Smith, F.D., Sullivan, K.M.E., and Franks, N.R. (2009). Do ants make direct comparisons? Proc. R. Soc. Lond. B Biol. Sci. *276*, 2635–2641.

Rodd, F.H., Hughes, K.A., Grether, G.F., and Baril, C.T. (2002). A possible non-sexual origin of mate preference: are male guppies mimicking fruit? Proc. R. Soc. Lond. B Biol. Sci. *269*, 475–481.

van Schaik, C.P., van Noordwijk, M.A., Warsono, B., and Sutriono, E. (1983). Party size and early detection of predators in Sumatran forest primates. Primates *24*, 211–221.

Seeley, T.D., and Visscher, P.K. (2004). Quorum sensing during nest-site selection by honeybee swarms. Behav. Ecol. Sociobiol. *56*, 594–601.

Seeley, T.D., Visscher, P.K., and Passino, K.M. (2006). Group decision making in honey bee swarms. Am. Sci. *94*, 220–229.

Shanteau, J., Weiss, D.J., Thomas, R.P., and Pounds, J.C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. Eur. J. Oper. Res. *136*, 253–263.

Sih, A., Bell, A., and Johnson, J.C. (2004a). Behavioral syndromes: an ecological and evolutionary overview. Trends Ecol. Evol. *19*, 372–378.

Sih, A., Bell, A.M., Johnson, J.C., and Ziemba, R.E. (2004b). Behavioral syndromes: an integrative overview. Q. Rev. Biol. *79*, 241–277.

Simons, A.M. (2004). Many wrongs: the advantage of group navigation. Trends Ecol. Evol. *19*, 453–455.

Sirot, E. (2006). Social information, antipredatory vigilance and flight in bird flocks. Anim. Behav. *72*, 373–382.

Sorkin, R.D., Hays, C.J., and West, R. (2001). Signal-detection analysis of group decision making. Psychol. Rev. *108*, 183–203.

Strandburg-Peshkin, A., Twomey, C.R., Bode, N.W.F., Kao, A.B., Katz, Y., Ioannou, C.C., Rosenthal, S.B., Torney, C.J., Wu, H.S., Levin, S.A., et al. (2013). Visual sensory networks and effective information transfer in animal groups. Curr. Biol. *23*, R709–R711.

Sueur, C., Deneubourg, J.-L., and Petit, O. (2010). Sequence of quorums during collective decision making in macaques. Behav. Ecol. Sociobiol. *64*, 1875–1885.

Sumpter, D.J.T. (2010). Collective Animal Behavior (Princeton, USA: Princeton University Press).

Sumpter, D.J.T., and Pratt, S.C. (2009). Quorum responses and consensus decision-making. Philos. Trans. R. Soc. B Biol. Sci. *364*, 743–753.

Sumpter, D.J.T., Krause, J., James, R., Couzin, I.D., and Ward, A.J.W. (2008a). Consensus decision making by fish. Curr. Biol. *18*, 1773–1777.

Sumpter, D.J.T., Buhl, J., Biro, D., and Couzin, I.D. (2008b). Information transfer in moving animal groups. Theory Biosci. *127*, 177–186.

Sunstein, C.R. (2005). Group judgments: Statistical means, deliberation, and information markets. N. Y. Univ. Law Rev. *80*, 962–1049.

Surowiecki, J. (2004). The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations (USA: Doubleday).

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. Science *240*, 1285–1293.

Swets, J.A., Dawes, R.M., and Monahan, J. (2000). Psychological science can improve diagnostic decisions. Psychol. Sci. Public Interest *1*, 1–26.

Templeton, J.J., and Giraldeau, L.-A. (1995). Patch assessment in foraging flocks of European starlings: evidence for the use of public information. Behav. Ecol. *6*, 65–72.

Tindale, R.S., and Larson, J.R. (1992). Assembly bonus effect or typical group performance? A comment on Michaelsen, Watson, and Black (1989). J. Appl. Psychol. *77*, 102–105.

Treherne, J.E., and Foster, W.A. (1980). The effects of group size on predator avoidance in a marine insect. Anim. Behav. *28*, 1119–1122.

Treherne, J.E., and Foster, W.A. (1981). Group transmission of predator avoidance behaviour in a marine insect: The Trafalgar effect. Anim. Behav. *29*, 911–917.

Visscher, P.K., and Camazine, S. (1999). Collective decisions and cognition in bees. Nature *397*, 400.

Wallraff, H.G. (1978). Social interrelations involved in migratory orientation of birds: possible contribution of field studies. Oikos *30*, 401–404.

Wallraff, H.G. (2001). Navigation by homing pigeons: updated perspective. Ethol. Ecol. Evol. *13*, 1–48.

Warburton, K. (2003). Learning of foraging skills by fish. Fish Fish. *4*, 203–215.

Ward, A.J.W., Sumpter, D.J.T., Couzin, I.D., Hart, P.J.B., and Krause, J. (2008). Quorum decision-making facilitates information transfer in fish shoals. Proc. Natl. Acad. Sci. U. S. A. *105*, 6948–6953.

Ward, A.J.W., Herbert-Read, J.E., Sumpter, D.J.T., and Krause, J. (2011). Fast and accurate decisions through collective vigilance in fish shoals. Proc. Natl. Acad. Sci. U. S. A. *108*, 2312–2315.

Ward, A.J.W., Krause, J., and Sumpter, D.J.T. (2012). Quorum decision-making in foraging fish shoals. PLoS One *7*, e32411.

Weber, M. (1978). The Nature of Social Action. In Selections in Translation, W.G. Runciman, ed. (Cambridge, U.K.: Cambridge University Press), pp. 7–32.

Webster, M.M., and Laland, K.N. (2008). Social learning strategies and predation risk: minnows copy only when using private information would be costly. Proc. R. Soc. Lond. B Biol. Sci. *275*, 2869–2876.

Webster, M.M., and Laland, K.N. (2013). The learning mechanism underlying public information use in ninespine sticklebacks (*Pungitius pungitius*). J. Comp. Psychol. *127*, 154.

Wolf, M., and Krause, J. (2014). Why personality differences matter for social functioning and social structure. Trends Ecol. Evol. *29*, 306–308.

Wolf, M., Kurvers, R.H.J.M., Ward, A.J.W., Krause, S., and Krause, J. (2013). Accurate decisions in an uncertain world: collective cognition increases true positives while decreasing false positives. Proc. R. Soc. B Biol. Sci. *280*, 1471–2954.

Wolf, M., Krause, J., Carney, P.A., Bogart, A., and Kurvers, R.H.J.M. (2015). Collective intelligence meets medical decision-making: the collective outperforms the best radiologist. PLoS One *10*, e0134269.

Wolfers, J., and Zitzewitz, E. (2004). Prediction markets. J. Econ. Perspect. *18*, 107–126.

Zweig, M.H., and Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. *39*, 561–577.